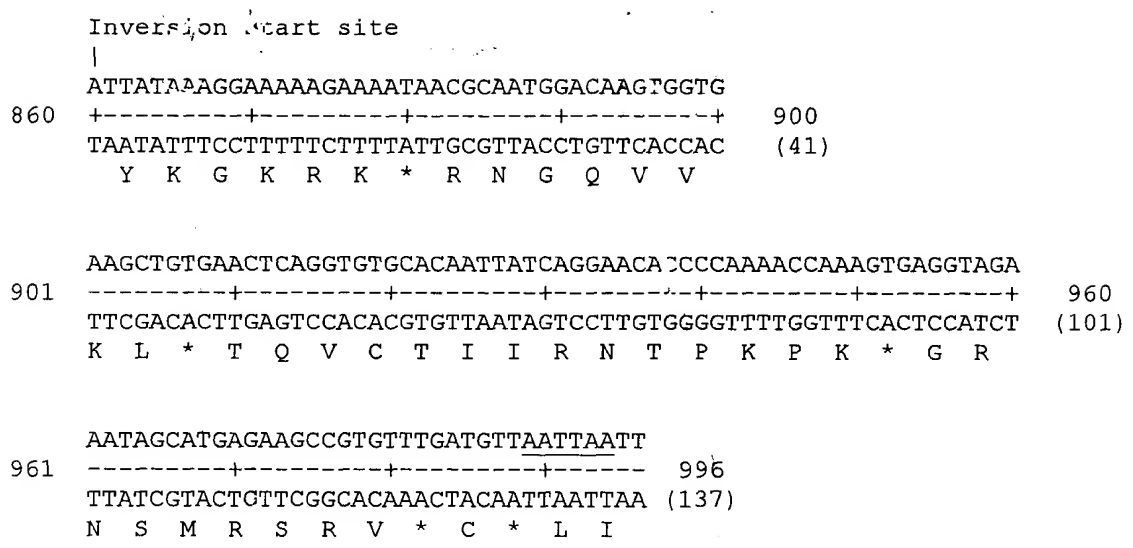


The inversion sequence of the apo-dystrophin-4 cDNA (SEQ ID NO 1)

Figure 1

```
                   Inversion start site
                    |
     TAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
850  +---------+---------+---------+---------+---------+   900
     ATTTCTTTCTTAATATTTCCTTTTTCTTTTATTGCGTTACCTGTTCACCAC   (51)
     *  R  K  N  Y  K  G  K  R  K  *  R  N  G  Q  V  V

     AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA
901  ---------+---------+---------+---------+---------+---------+   960
     TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTTGGTTTCACTCCATCT   (111)
     K  L  *  T  Q  V  C  T  I  I  R  N  T  P  K  P  K  *  G  R

     AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT
961  ---------+---------+---------+------   996
     TTATCGTACTCTTCGGCACAAACTACAATTAATTAA   (147)
     N  S  M  R  S  R  V  *  C  *  L  I
```

The inversion sequence of the apo-dystrophin-4 cDNA plus a 10 base-pair region 5' to the start of the inversion sequence (SEQ ID NO 1A).

Figure 1A

```
     Start at 710
     |
     AACAATGGCAG
     +---------+   720
     TTGTTACCGTC   (11)
       Q  W  Q

     GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG
721  ---------+---------+---------+---------+---------+---------+   780
     CAAAATGTGCAGATACGTTAACATGTTTTTTCAATATTCTTTTGATGTACATTTTAGAAC   (71)
     V  L  H  V  Y  A  I  V  Q  K  S  Y  K  K  T  T  C  K  I  L

     ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA
781  ---------+---------+---------+---------+---------+---------+   840
     TATCGATTTATTGAACGGTAAAGAAATATACCTTGCGTAAAACCCAACAAATTTTTAAAT   (131)
     I  A  K  *  L  A  I  S  L  Y  G  T  H  F  G  L  F  K  N  L
                  inversion start site
                                 |
     TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
841  ---------+---------+---------+---------+---------+---------+   900
     ATTGTCAATATTTCTTTCTTAATATTTCCTTTTTCTTTTATTGCGTTACCTGTTCACCAC   (191)
     *  Q  L  *  R  K  N  Y  K  G  K  R  K  *  R  N  G  Q  V  V

     AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA
901  ---------+---------+---------+---------+---------+---------+   960
     TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTTGGTTTCACTCCATCT   (251)
     K  L  *  T  Q  V  C  T  I  I  R  N  T  P  K  P  K  *  G  R

     AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT
961  ---------+---------+---------+------   996
     TTATCGTACTCTTCGGCACAAACTACAATTAATTAA  (287)
     N  S  M  R  S  R  V  *  C  *  L  I
```

The inversion sequence of the apo-dystrophin-4 cDNA plus the upstream 150 bp from the start of the inversion at 860 to the Hpa I enzyme site (SEQ ID NO 1B)

Figure 1B

```
    GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT
  1 ---------+---------+---------+---------+---------+---------+ 60
    CACCAAACTAACTATCATTTTTTTTACAAGCAATTATGTTCATCTCTCATTCATTAGTTA
    V  V  *  L  I  V  K  K  M  F  V  N  T  S  R  E  *  V  I  N

    CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT
 61 ---------+---------+---------+---------+---------+---------+ 120
    GTTAGTGAGTATCGGTTCCACCTTTTCTACATAGGGTAGTACCTTATAAGGACAAGACTA
    Q  S  L  I  A  K  V  E  K  M  Y  P  I  M  E  Y  S  C  S  D

    AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT
121 ---------+---------+---------+---------+---------+---------+ 180
    TCTTTAGAACACGAATAGATACCTTAAGAAAACTATATATAAATGTAACCCTTGGACTTA
    R  N  L  V  L  I  Y  G  I  L  L  I  Y  I  Y  I  G  N  L  N

    GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA
181 ---------+---------+---------+---------+---------+---------+ 240
    CATCGAACTGTAAAAAGGTACATTTGTGGTCATCGGACTAGGTTGTAATTCGACTATGAT
    V  A  *  H  F  S  M  *  T  P  V  A  *  S  N  I  K  L  I  L

    ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA
241 ---------+---------+---------+---------+---------+---------+ 300
    TGTTTGTTGCACATTACCGAAGTAATTATTCCGAAACGAAGAAGGACCTTTGACCACTTT
    T  N  N  V  *  W  L  H  *  *  G  F  A  S  S  W  K  L  V  K

    AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG
301 ---------+---------+---------+---------+---------+---------+ 360
    TTAGTTTGGAACAACACATGTGGGAGCTACGTCGAAGACACAACAGAAGTGGGTCTTTAC
    N  Q  T  L  L  C  T  P  S  M  Q  L  L  C  C  L  H  P  E  M
```

The polynucleotide sequence of apo-dystrophin-4 (SEQ ID NO 2)

Figure 2

```
    GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA
361 ---------+---------+---------+---------+---------+---------+ 420
    CCCTTACTAAAGGGTTTACCGTTTCTTTGTCTCACTACGATAGATAGACGTGGAAAACAT
    G  N  D  F  P  N  G  K  E  T  E  *  C  Y  L  S  A  P  F  V

    AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG
421 ---------+---------+---------+---------+---------+---------+ 480
    TTCAGACAGAAAGAAAGAGAAACAAAAGGTCCTGTGTTACATCCTTCAGAAAAGGTGTAC
    K  S  V  F  L  S  L  C  F  P  G  H  N  V  G  S  L  F  H  M

    GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA
481 ---------+---------+---------+---------+---------+---------+ 540
    CGTCTACTAAACCCGTCTCGCTACCTCAGGAATCATAGTCAGTACTGTCTACTTCTTCCT
    A  D  D  L  G  R  A  M  E  S  L  V  S  V  M  T  D  E  E  G

    GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA
541 ---------+---------+---------+---------+---------+---------+ 600
    CGTCTTATTTACAAAATGTTGAGGACTAAGGGCGTACCAAAAATATTATAAGTATGTTGT
    A  E  *  M  F  Y  N  S  *  F  P  H  G  F  Y  N  I  H  T  T

    AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG
601 ---------+---------+---------+---------+---------+---------+ 660
    TTCTCCTAATCTGTCATTCTCAAATGTTCTTTATTTAGATATAAAAACACTTCCCATCAC
    K  R  I  R  Q  *  E  F  T  R  N  K  S  I  F  L  *  R  V  V

    GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG
661 ---------+---------+---------+---------+---------+---------+ 720
    CATAATATGACATCTAAAGTCATCAAAGATTCAGACAATAACAAAACAATTGTTACCGTC
    V  L  Y  C  R  F  Q  *  F  L  S  L  L  L  F  C  *  Q  W  Q
```


Figure 2 (cont'd)

```
     GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG
721  ---------+---------+---------+---------+---------+---------+  780
     CAAAATGTGCAGATACGTTAACATGTTTTTTCAATATTCTTTTGATGTACATTTTAGAAC
     V  L  H  V  Y  A  I  V  Q  K  S  Y  K  K  T  T  C  K  I  L

     ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA
781  ---------+---------+---------+---------+---------+---------+  840
     TATCGATTTATTGAACGGTAAAGAAATATACCTTGCGTAAAACCCAACAAATTTTTAAAT
     I  A  K  *  L  A  I  S  L  Y  G  T  H  F  G  L  F  K  N  L

     TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
841  ---------+---------+---------+---------+---------+---------+  900
     ATTGTCAATATTTCTTTCTTAATATTTCCTTTTTCTTTTATTGCGTTACCTGTTCACCAC
     *  Q  L  *  R  K  N  Y  K  G  K  R  K  *  R  N  G  Q  V  V

     AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA
901  ---------+---------+---------+---------+---------+---------+  960
     TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTTGGTTTCACTCCATCT
     K  L  *  T  Q  V  C  T  I  I  R  N  T  P  K  P  K  *  G  R

     AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT
961  ---------+---------+---------+------  996
     TTATCGTACTCTTCGGCACAAACTACAATTAATTAA
     N  S  M  R  S  R  V  *  C  *  L  I
```

Figure 2 (cont'd)

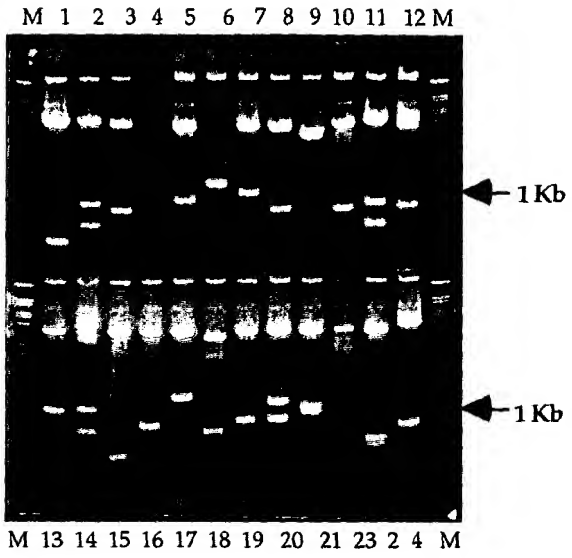


Figure 3A

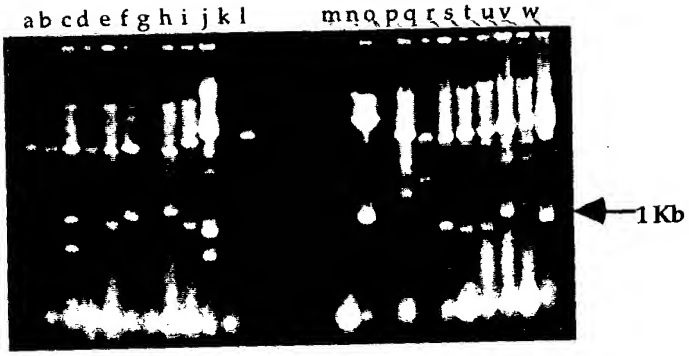


Figure 3B

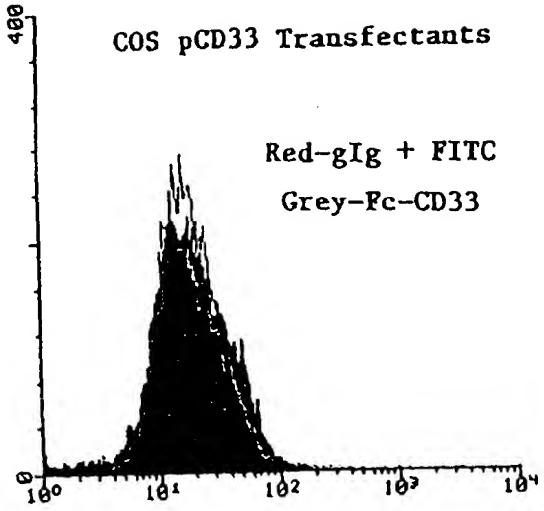


Figure 4A

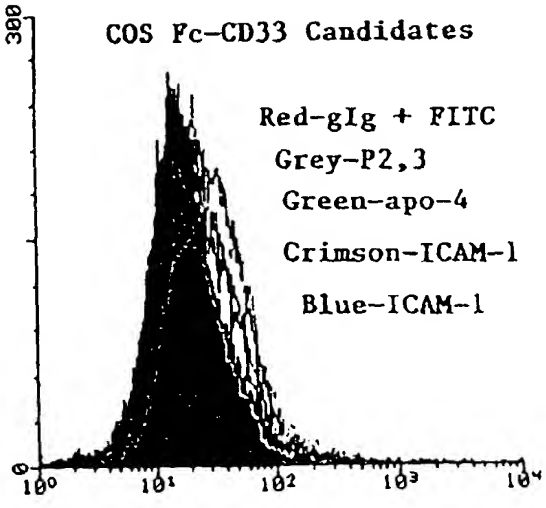


Figure 4B

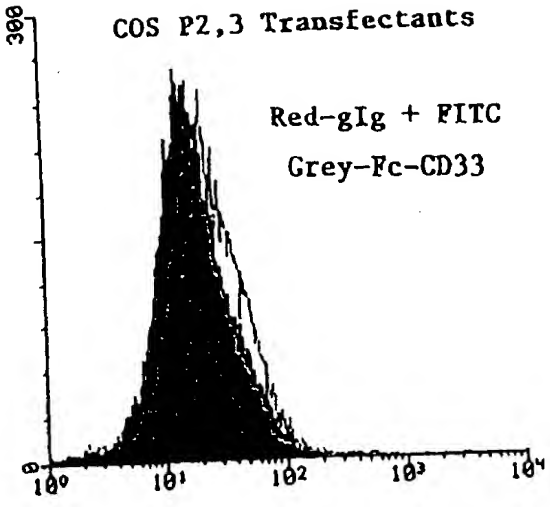
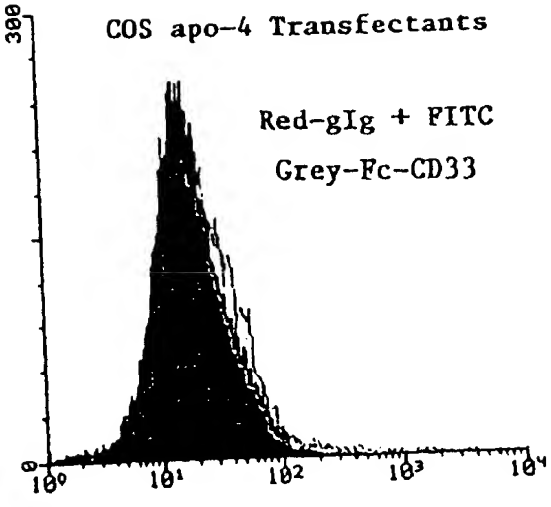


Figure 4C



Figur 4D

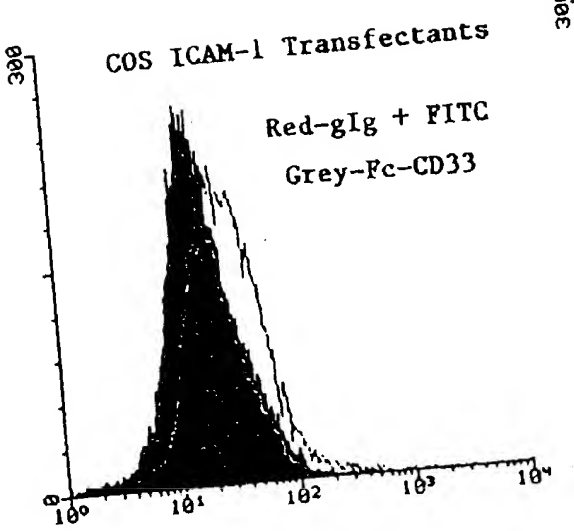


Figure 4E

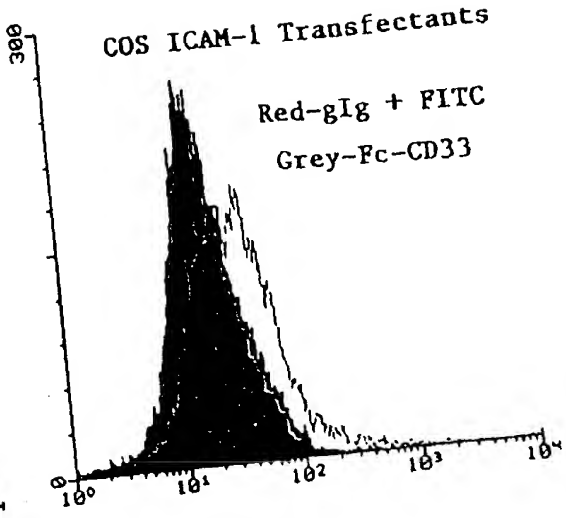


Figure 4F

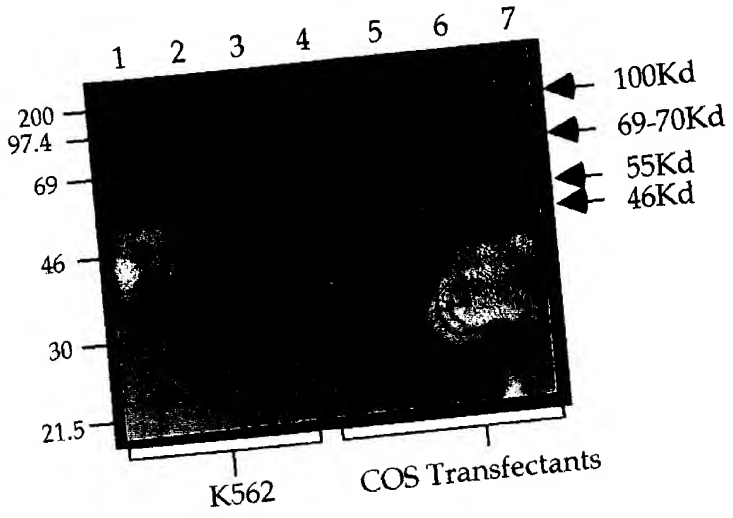


Figure 5

```
       ...TAGTTTCCTATTCAATGTATAGTGCACCAAAGGTCAATTCAAGAGTTTATTATTATT
-239  ---------+---------+---------+---------+---------+---------+  -180
       ...ATCAAAGGATAAGTTACATATCACGTGGTTTCCAGTTAAGTTCTCAAATAATAATAA
       .  *  F  P  I  Q  C  I  V  H  Q  R  S  I  Q  E  F  I  I  I

      ATTTTCAACCCAAGTAAAAGCAGAGAGAAAATAGCCACCTCCACCATAGCCTCAGAAGCA
-179  ---------+---------+---------+---------+---------+---------+  -120
      TAAAAGTTGGGTTCATTTTCGTCTCTCTTTTATCGGTGGAGGTGGTATCGGAGTCTTCGT
      I  F  N  P  S  K  S  R  E  K  I  A  T  S  T  I  A  S  E  A

      AGCCAACAGCCTGAAACAGCTTTGAAATGAAAAGTTGGTGTGGCGGTGATGGTGGCAGTG
-119  ---------+---------+---------+---------+---------+---------+  -60
      TCGGTTGTCGGACTTTGTCGAAACTTTACTTTTCAACCACACCGCCACTACCACCGTCAC
      S  Q  Q  P  E  T  A  L  K  *  K  V  G  V  A  V  M  V  A  V

      ATAATGGTGACCGATGGTTGGGTGCTGGTGATGGTAGTGGTAGTTGTGAAGGTGGTGATG
-59   ---------+---------+---------+---------+---------+---------+  0
      TATTACCACTGGCTACCAACCCACGACCACTACCATCACCATCAACACTTCCACCACTAC
      I  M  V  T  D  G  W  V  L  V  M  V  V  V  V  V  K  V  V  M

      GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT
1     ---------+---------+---------+---------+---------+---------+  60
      CACCAAACTAACTATCATTTTTTTTACAAGCAATTATGTTCATCTCTCATTCATTAGTTA
      V  V  *  L  I  V  K  K  M  F  V  N  T  S  R  E  *  V  I  N

      CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT
61    ---------+---------+---------+---------+---------+---------+  120
      GTTAGTGAGTATCGGTTCCACCTTTTCTACATAGGGTAGTACCTTATAAGGACAAGACTA
      Q  S  L  I  A  K  V  E  K  M  Y  P  I  M  E  Y  S  C  S  D

      AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT
121   ---------+---------+---------+---------+---------+---------+  180
      TCTTTAGAACACGAATAGATACCTTAAGAAAACTATATATAAATGTAACCCTTGGACTTA
      R  N  L  V  L  I  Y  G  I  L  L  I  Y  I  Y  I  G  N  L  N

      GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA
181   ---------+---------+---------+---------+---------+---------+  240
      CATCGAACTGTAAAAAGGTACATTTGTGGTCATCGGACTAGGTTGTAATTCGACTATGAT
      V  A  *  H  F  S  M  *  T  P  V  A  *  S  N  I  K  L  I  L

      ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA
241   ---------+---------+---------+---------+---------+---------+  300
      TGTTTGTTGCACATTACCGAAGTAATTATTCCGAAACGAAGAAGGACCTTTGACCACTTT
      T  N  N  V  *  W  L  H  *  *  G  F  A  S  S  W  K  L  V  K

      AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG
301   ---------+---------+---------+---------+---------+---------+  360
      TTAGTTTGGAACAACACATGTGGGAGCTACGTCGAAGACACAACAGAAGTGGGTCTTTAC
      N  Q  T  L  L  C  T  P  S  M  Q  L  L  C  C  L  H  P  E  M

      GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA
361   ---------+---------+---------+---------+---------+---------+  420
      CCCTTACTAAAGGGTTTACCGTTTCTTTGTCTCACTACGATAGATAGACGTGGAAAACAT
      G  N  D  F  P  N  G  K  E  T  E  *  C  Y  L  S  A  P  F  V
```

Figure 6

```
                              begin exon 79
                                    |
     AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG
421  ---------+---------+---------+---------+---------+---------+  480
     TTCAGACAGAAAGAAAGAGAAACAAAAGGTCCTGTGTTACATCCTTCAGAAAAGGTGTAC
     K  S  V  F  L  S  L  C  F  P  G  H  N  V  G  S  L  F  H  M

     GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA
481  ---------+---------+---------+---------+---------+---------+  540
     CGTCTACTAAACCCGTCTCGCTACCTCAGGAATCATAGTCAGTACTGTCTACTTCTTCCT
     A  D  D  L  G  R  A  M  E  S  L  V  S  V  M  T  D  E  E  G

     GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA
541  ---------+---------+---------+---------+---------+---------+  600
     CGTCTTATTTACAAAATGTTGAGGACTAAGGGCGTACCAAAAATATTATAAGTATGTTGT
     A  E  *  M  F  Y  N  S  *  F  P  H  G  F  Y  N  I  H  T  T

                                (----N----)
     AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG
601  ---------+---------+---------+---------+---------+---------+  660
     TTCTCCTAATCTGTCATTCTCAAATGTTCTTTATTTAGATATAAAAACACTTCCCATCAC
     K  R  I  R  Q  *  E  F  T  R  N  K  S  I  F  L  *  R  V  V

     GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG
661  ---------+---------+---------+---------+---------+---------+  720
     CATAATATGACATCTAAAGTCATCAAAGATTCAGACAATAACAAAACAATTGTTACCGTC
     V  L  Y  C  R  F  Q  *  F  L  S  L  L  L  F  C  *  Q  W  Q

     GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG
721  ---------+---------+---------+---------+---------+---------+  780
     CAAAATGTGCAGATACGTTAACATGTTTTTTCAATATTCTTTTGATGTACATTTTAGAAC
     V  L  H  V  Y  A  I  V  Q  K  S  Y  K  K  T  T  C  K  I  L

     ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA
781  ---------+---------+---------+---------+---------+---------+  840
     TATCGATTTATTGAACGGTAAAGAAATATACCTTGCGTAAAACCCAACAAATTTTTAAAT
     I  A  K  *  L  A  I  S  L  Y  G  T  H  F  G  L  F  K  N  L
                inversion start site
                         |
     TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
841  ---------+---------+---------+---------+---------+---------+  900
     ATTGTCAATATTTCTTTCTTAATATTTCCTTTTTCTTTTATTGCGTTACCTGTTCACCAC
     *  Q  L  *  R  K  N  Y  K  G  K  R  K  *  R  N  G  Q  V  V

     AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA
901  ---------+---------+---------+---------+---------+---------+  960
     TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTTGGTTTCACTCCATCT
     K  L  *  T  Q  V  C  T  I  I  R  N  T  P  K  P  K  *  G  R

     AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT
961  ---------+---------+---------+------  996
     TTATCGTACTCTTCGGCACAAACTACAATTAATTAA
     N  S  M  R  S  R  V  *  C  *  L  I
```

Figure 6 (cont'd)

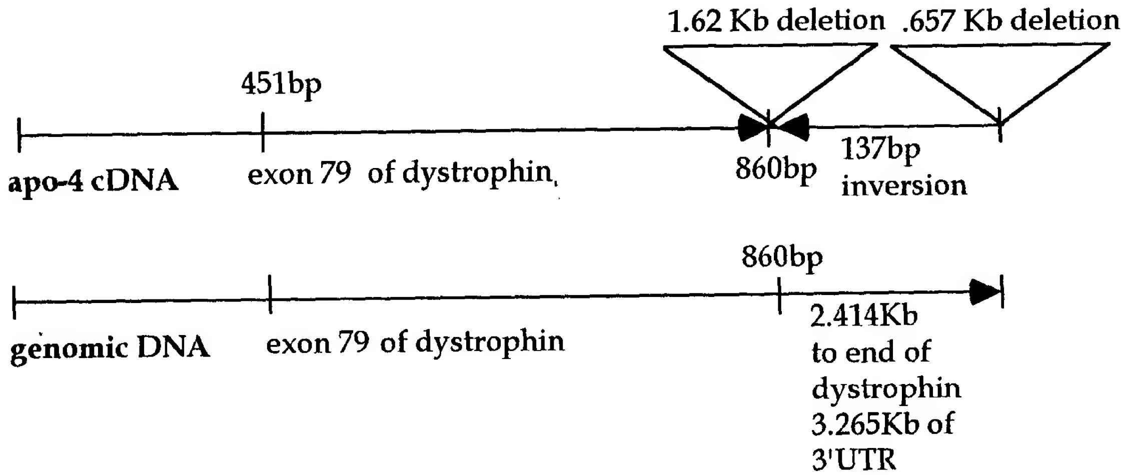
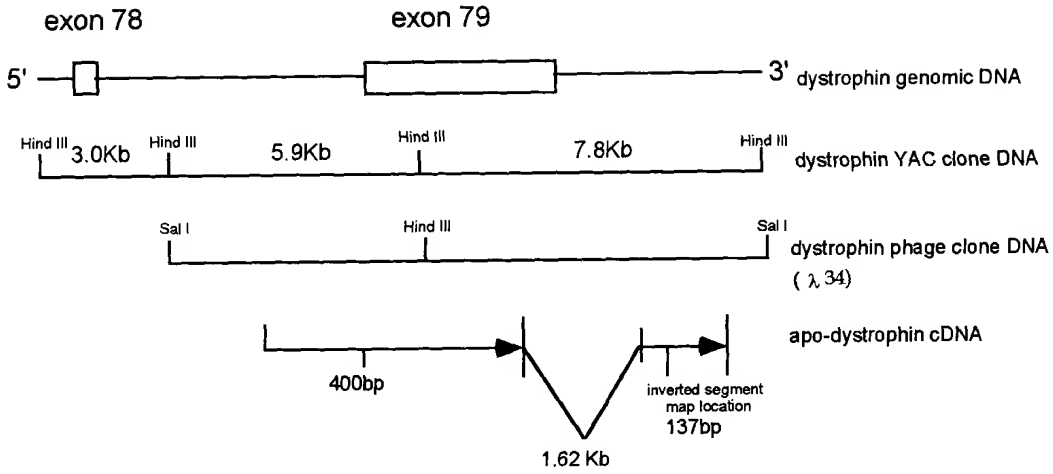
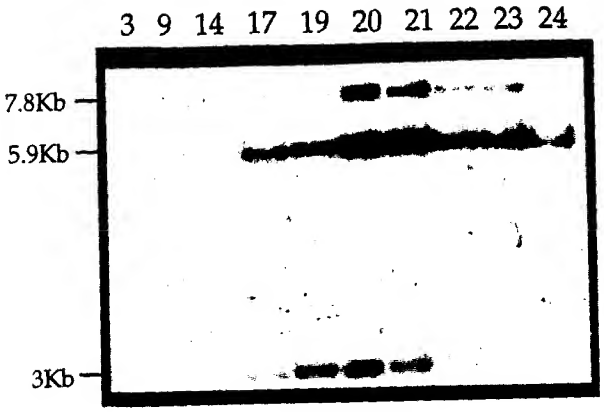

1.62 Kb deletion
.657 Kb deletion
451bp
apo-4 cDNA
exon 79 of dystrophin
860bp
137bp
inversion
860bp
genomic DNA
exon 79 of dystrophin
2.414Kb
to end of
dystrophin
3.265Kb of
3'UTR

Figure 7



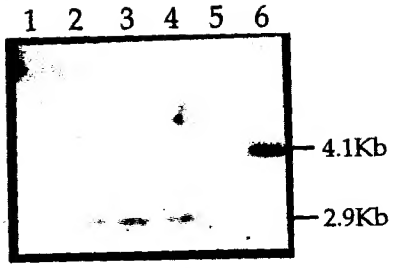
*cDNA map is not precisely drawn to scale

Figure 8



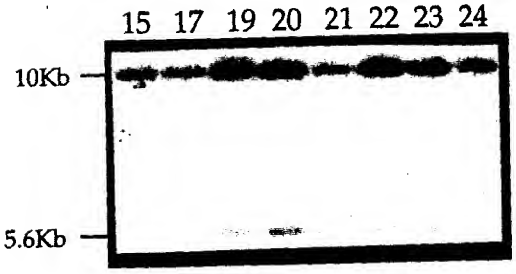
A. Human

Figure 9A



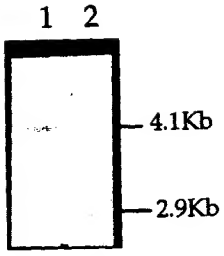
B. Mouse

Figure 9B



A. Human

Figure 10A



B. Mouse

Figure 10B

```
          50                                                        1
Mgen1073  .......... .......... .......... .......... ..........
Hapo1234  ctagtttcct attcaatgta tagtgcacca aaggtcaatt caagagttta
Consensus ---------- ---------- ---------- ---------- ----------

          51                                                      100
Mgen1073  .......... .......... .......... .......... ..........
Hapo1234  ttattattat tttcaaccca agtaaaagca gagagaaaat agccacctcc
Consensus ---------- ---------- ---------- ---------- ----------

          101                          begin GRAIL exon @149 150
Mgen1073  .......... .......... ttcACAGgCT tAAgCAGCca gtAAATGAcA
Hapo1234  accatagcct cagaagcaag ccaACAGcCT gAAaCAGCtt tgAAATGAaA
Consensus ---------- ---------- ---ACAG-CT -AA-CAGC-- --AAATGA-A

          151                                                     200
Mgen1073  AtT....... .......... .........T AtgtGgtAgt cAgGtcactG
Hapo1234  AgTtggtgtg gcggtgatgg tggcagtgaT AatgGtgAcc gAtGgttggG
Consensus A-T------- ---------- ---------T A---G--A-- -A-G-----G

          201                                    apo-4 5'end 250
Mgen1073  TGCTGGTaAT GGTgaTctTA GcaGgcAgAG aaGGTGgTaG TGaTTTGATa
Hapo1234  TGCTGGTgAT GGTagTggTA GttGtgA.AG gtGGTGaTgG TGgTTTGATt
Consensus TGCTGGT-AT GGT--T--TA G--G--A-AG --GGTG-T-G TG-TTTGAT-

          251          M1                                         300
Mgen1073  GtaAaagtgt AgAcTaTaCa acAgaAtAAa TAcAagtatA GTAA......
Hapo1234  GatAgtaaaa AaAaTgTtCg ttAatAcAAg TAgAgagtaA GTAAtcaatc
Consensus G--A------ A-A-T-T-C- --A--A-AA- TA-A-----A GTAA------

          301                          M2          M3      350
Mgen1073  .......atc caaCAAaGTG tgAAAGgTGT gTgCCATtAc acAtctTTCt
Hapo1234  aatcactcat agcCAAgGTG gaAAAGaTGT aTcCCATcAt ggAataTTCc
Consensus ---------- ---CAA-GTG --AAAG-TGT -T-CCAT-A- --A---TTC-

          351                                                     400
Mgen1073  cG........ GtgATaagag cCTTgTCTAT GaAgTTC... TGAgATgTgT
Hapo1234  tGttctgata GaaATcttgt gCTTaTCTAT GgAaTTCttt TGAtATaTaT
Consensus -G-------- G--AT----- -CTT-TCTAT G-A-TTC--- TGA-AT-T-T

          401                                                     450
Mgen1073  TaggAagatG AAtCatcAat TtaCaT.... TTcTcCCcat cAAAtgaCAc
Hapo1234  TtacAttggG AAcCtgaAtg TagCtTgaca TTtTtCCatg tAAAcacCAg
Consensus T---A----G AA-C---A-- T--C-T---- TT-T-CC--- -AAA---CA-

          451              begin mouse GRAIL exon                 500
Mgen1073  cAtgCTGATC CAgtATTAAG CTaATACTAA C.....ACca tgcAatGCTT
Hapo1234  tAgcCTGATC CAacATTAAG CTgATACTAA CaaacaACgt gtaAtgGCTT
Consensus -A--CTGATC CA--ATTAAG CT-ATACTAA C-----AC-- ---A--GCTT

          501                                                     550
Mgen1073  CATTAAcAAG GaTTTGCTTC TTgCTaGAAA tgGGT..AAA AaCggACtgT
Hapo1234  CATTAAtAAG GcTTTGCTTC TTcCTgGAAA ctGGTgaAAA AtCaaACctT
Consensus CATTAA-AAG G-TTTGCTTC TT-CT-GAAA --GGT--AAA A-C--AC--T

          551                                                     600
Mgen1073  GgTcTGTAtA CCtTCaATGC AGCTTaTGTG TTGTCTTttC C..tgAAatG
Hapo1234  GtTgTGTAcA CCcTCgATGC AGCTTcTGTG TTGTCTTcaC CcagaAAtgG
Consensus G-T-TGTA-A CC-TC-ATGC AGCTT-TGTG TTGTCTT--C C----AA--G
```

Figure 11

```
           601                                                650
 Mgen1073  GtAATGAcTc CCaAtAgtGg cAAccAgggG tacaATaCT. ......TGCA
 Hapo1234  GgAATGAtTt CCcAaAtgGc aAAgaAacaG agtgATgCTa tctatcTGCA
Consensus  G-AATGA-T- CC-A-A--G- -AA--A---G ----AT-CT- ------TGCA

           651                                         exon79 700
 Mgen1073  CacTTTGTAA A....cTCTT TCTTTCTCTT TGTTTTCCAG GACACAATGT
 Hapo1234  CctTTTGTAA AgtctgTCTT TCTTTCTCTT TGTTTTCCAG GACACAATGT
Consensus  C--TTTGTAA A-----TCTT TCTTTCTCTT TGTTTTCCAG GACACAATGT

           701                                                750
 Mgen1073  AGGAAGcCTT TTCCACATGG CAGATGATTT GGGCAGAGCG ATGGAGTCCT
 Hapo1234  AGGAAGtCTT TTCCACATGG CAGATGATTT GGGCAGAGCG ATGGAGTCCT
Consensus  AGGAAG-CTT TTCCACATGG CAGATGATTT GGGCAGAGCG ATGGAGTCCT

           751                                                800
 Mgen1073  TAGTtTCAGT CATGACAGAT GAAGAAGGAG CAGAATAAAT GTTTTACAAC
 Hapo1234  TAGTaTCAGT CATGACAGAT GAAGAAGGAG CAGAATAAAT GTTTTACAAC
Consensus  TAGT-TCAGT CATGACAGAT GAAGAAGGAG CAGAATAAAT GTTTTACAAC

           801                                                850
 Mgen1073  TCCTGATTCC CGCATGGTTT TTATAATATT CgTACAACAA AGAGGATTAG
 Hapo1234  TCCTGATTCC CGCATGGTTT TTATAATATT CaTACAACAA AGAGGATTAG
Consensus  TCCTGATTCC CGCATGGTTT TTATAATATT C-TACAACAA AGAGGATTAG

           851                                                900
 Mgen1073  ACAGTAAGAG TTTACAAGAA ATaAAATCTA TATTTTTGTG AAGGGTAGTG
 Hapo1234  ACAGTAAGAG TTTACAAGAA AT.AAATCTA TATTTTTGTG AAGGGTAGTG
Consensus  ACAGTAAGAG TTTACAAGAA AT-AAATCTA TATTTTTGTG AAGGGTAGTG

           901                                                950
 Mgen1073  GTAcTATACT GTAGATTTCA GTAGTTTCTA AGTCTGTTAT TGTTTTGTTA
 Hapo1234  GTAtTATACT GTAGATTTCA GTAGTTTCTA AGTCTGTTAT TGTTTTGTTA
Consensus  GTA-TATACT GTAGATTTCA GTAGTTTCTA AGTCTGTTAT TGTTTTGTTA

           951                                               1000
 Mgen1073  ACAATGGCAG GTTTTACACG TCTATGCAAT TGTACAAAAA AGTTAaAAGA
 Hapo1234  ACAATGGCAG GTTTTACACG TCTATGCAAT TGTACAAAAA AGTTAtAAGA
Consensus  ACAATGGCAG GTTTTACACG TCTATGCAAT TGTACAAAAA AGTTA-AAGA

           1001                                              1050
 Mgen1073  AA...ACATG TAAAATCTTG ATAGCTAAAT AACTTGCCAT TTCTTTATAT
 Hapo1234  AActACATG  TAAAATCTTG ATAGCTAAAT AACTTGCCAT TTCTTTATAT
Consensus  AA---ACATG TAAAATCTTG ATAGCTAAAT AACTTGCCAT TTCTTTATAT
                                  begin inversion@1100
           1051                                              1100
 Mgen1073  GGAACGCATT TTGGGTTGTT TAAAAATTTA TAACAGTTAT AAAGAAAGAt
 Hapo1234  GGAACGCATT TTGGGTTGTT TAAAAATTTA TAACAGTTAT AAAGAAAGAa
Consensus  GGAACGCATT TTGGGTTGTT TAAAAATTTA TAACAGTTAT AAAGAAAGA-

           1101                                              1150
 Mgen1073  TgtaAActaA Agtgtgcttt AtAAAAaAAg ttgtTtataA AaacccctAa
 Hapo1234  TtatAAaggA A.......aa AgAAAAtAAc gcaaTggacA AgtggtgaAg
Consensus  T---AA---A A--------- A-AAAA-AA- ----T----A A-------A-

           1151                                              1200
 Mgen1073  acaaacACaC AcGcacaCAC AcacAcacac AcacaCaCAc AcaCAcAcTG
 Hapo1234  ctgtgaACtC AgGtgtgCAC AattAtcagg AacacCcCAa AacCAaAgTG
Consensus  ------AC-C A-G----CAC A---A----- A----C-CA- A--CA-A-TG

           1201                                   1243
 Mgen1073  AGGcAGcAca ttgtTttGcA ttacTtTagc gTGTatcaTA t..
 Hapo1234  AGGtAGaAat agcaTgaGaA gccgTgTttg aTGTtaatTA att
Consensus  AGG-AG-A-- ----T--G-A ----T-T--- -TGT----TA ---
```

Figure 11 (cont'd)

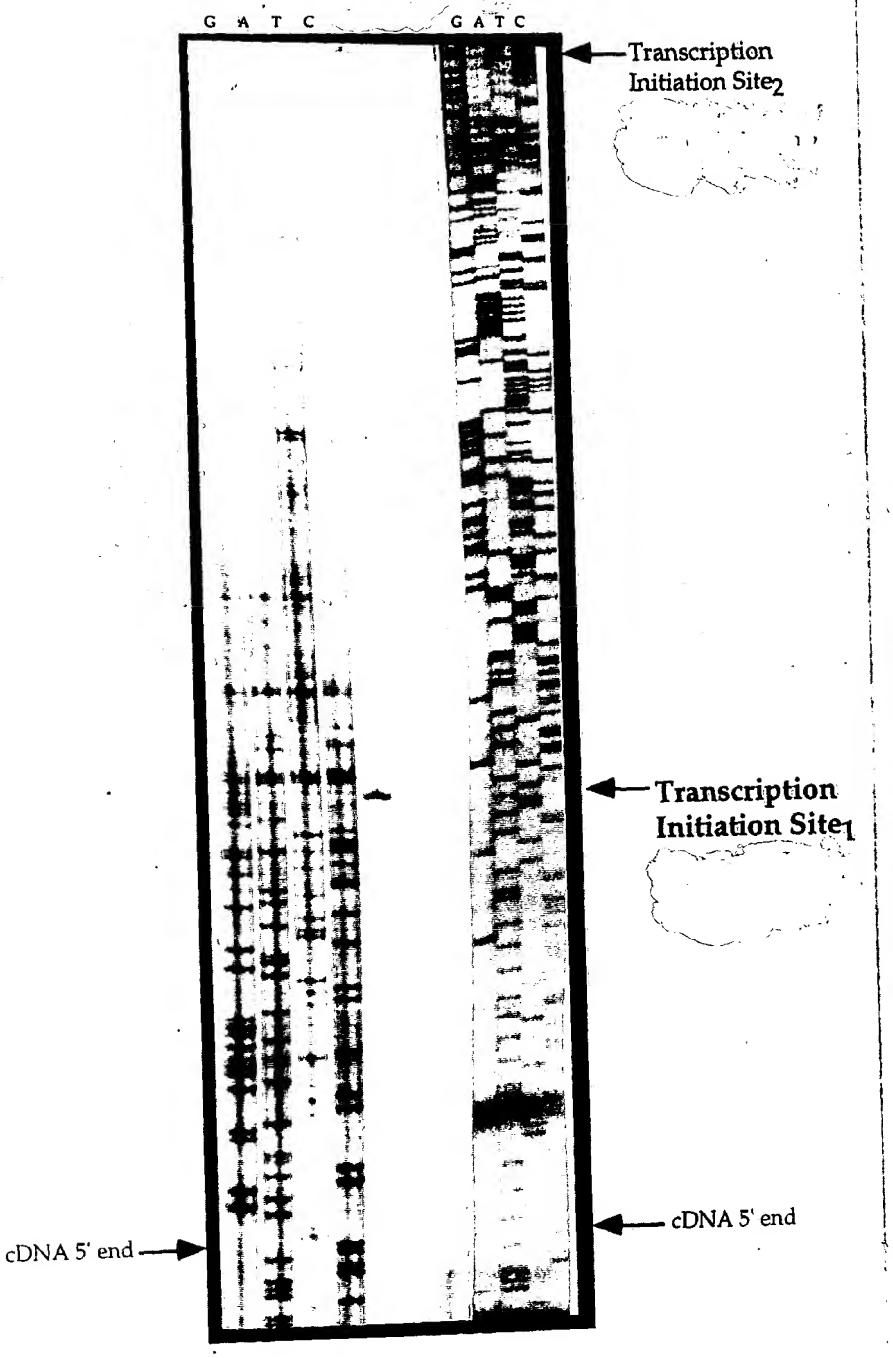

G A T C
G A T C
Transcription Initiation Site2
Transcription Initiation Site1
cDNA 5' end
cDNA 5' end

Figure 12A

-70 bp from 5' end of apo-4

```
Inr  = GCCC TCAT TCTG GAGAC
apo-4 = GCGG TGAT GGTG GCAGT
```

- 48% perfect homology with Inr
71% match on type of base
(purine vs. pyrimidine)

Figure 12B

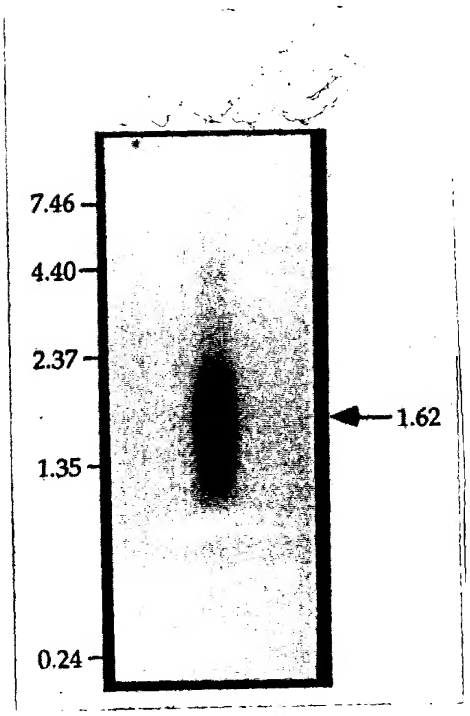


Figure 13

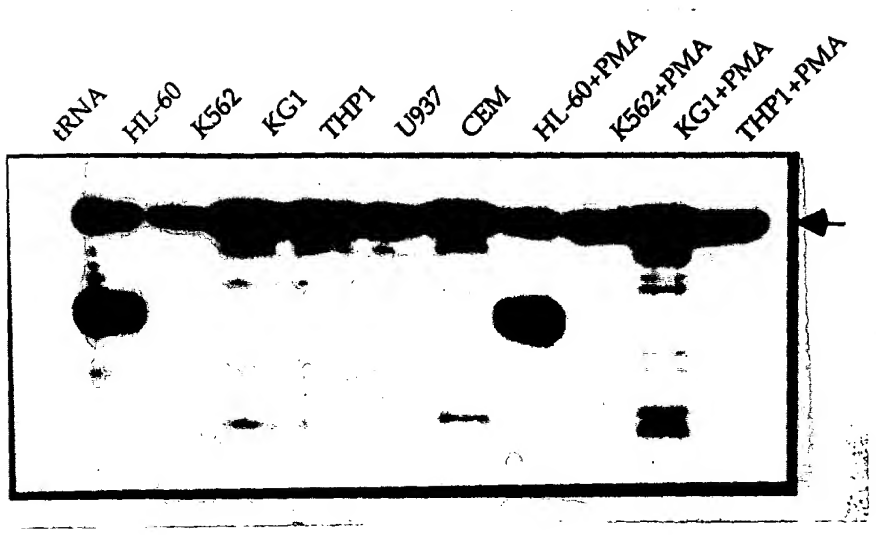


Figure 14

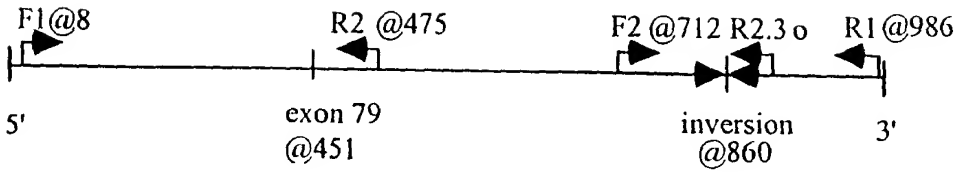
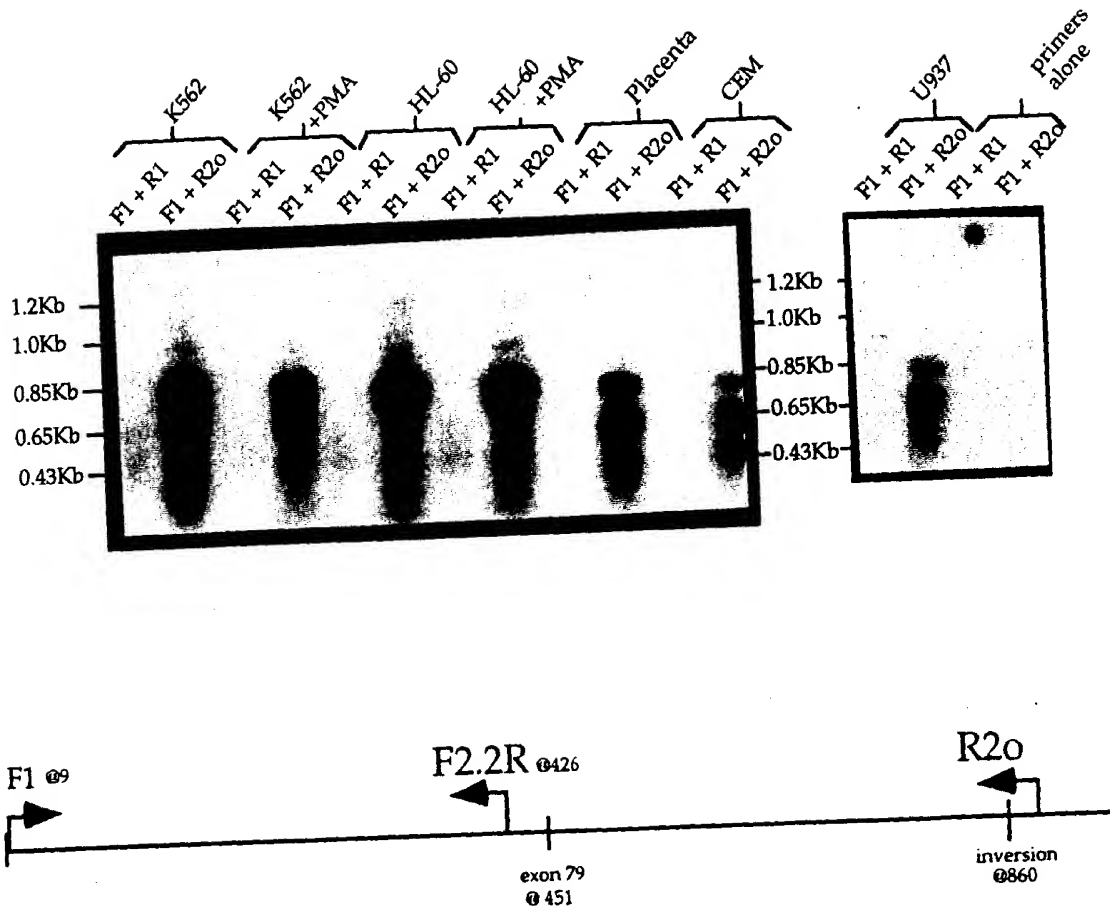


Figure 15



The Apo-dystrophin cDNA

Figure 16

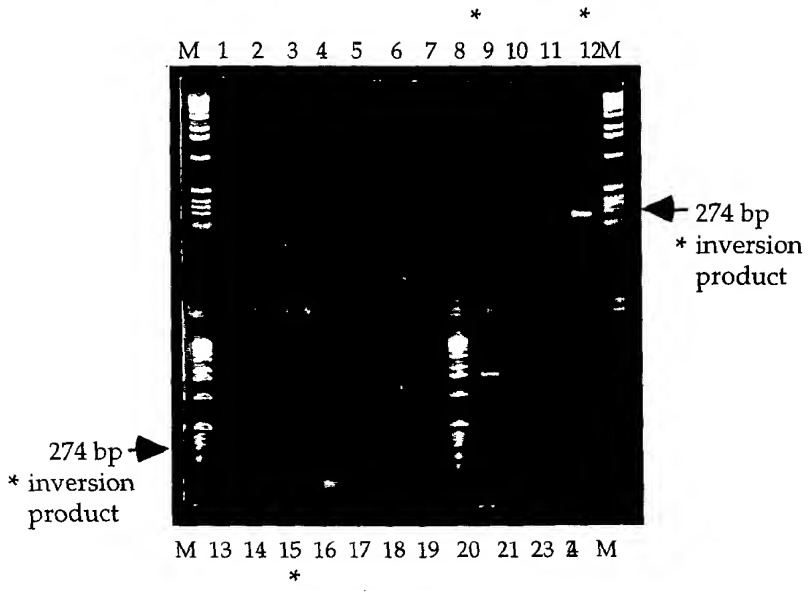
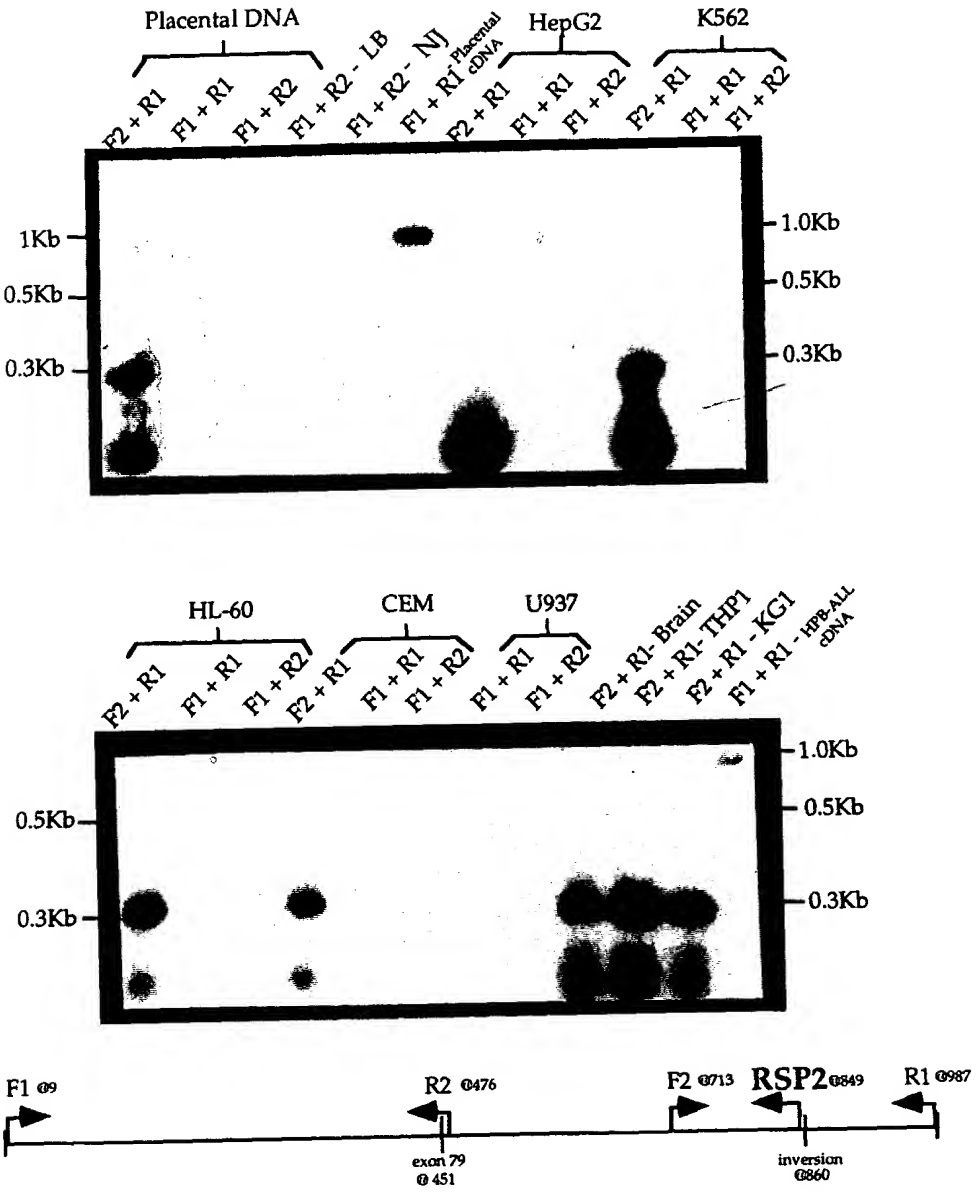
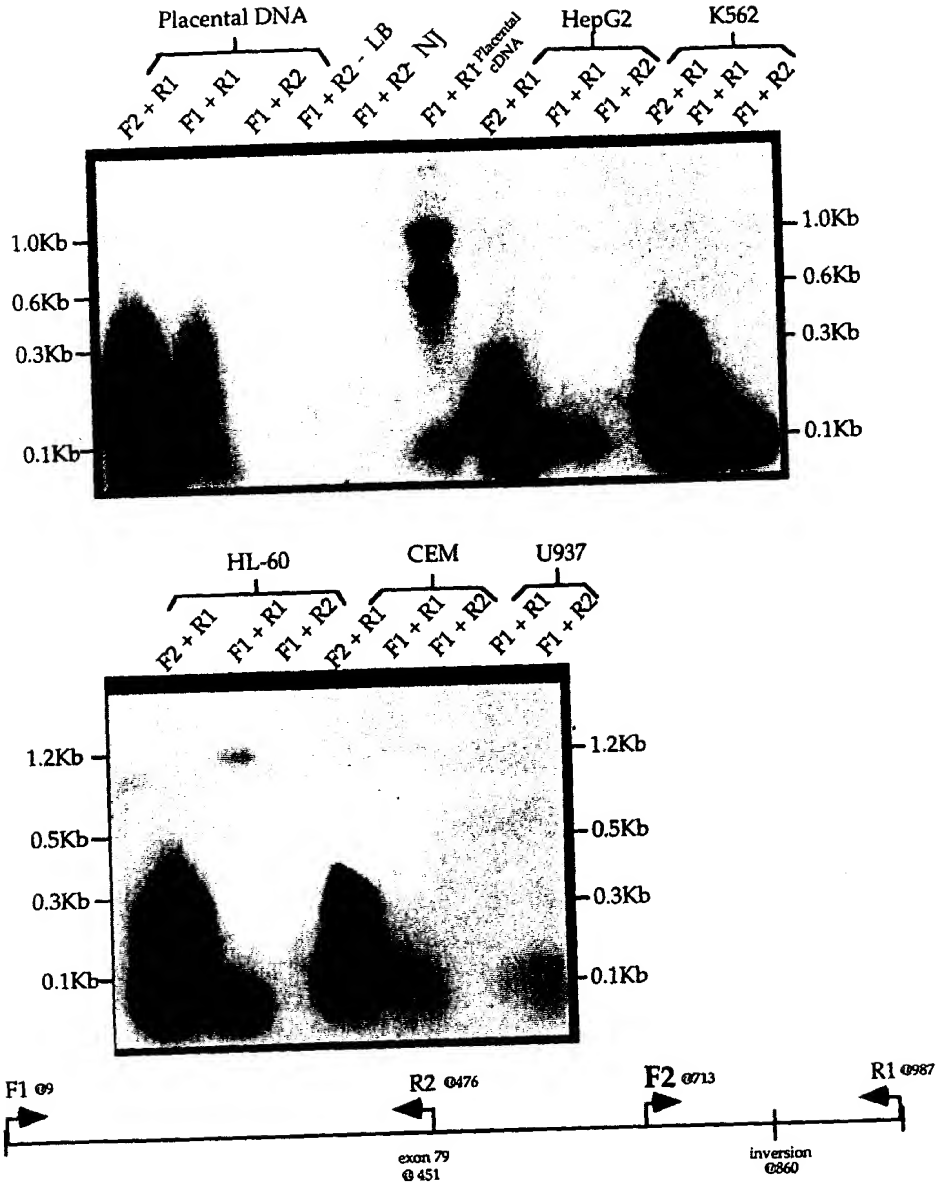

M 1 2 3 4 5 6 7 8 9 10 11 12M
274 bp
* inversion
product
274 bp
* inversion
product
M 13 14 15 16 17 18 19 20 21 23 24 M

Figure 17A



The Apo-dystrophin cDNA

Figure 17B



The Apo-dystrophin cDNA

Figure 17C

A.

```
                        12/23bp spacer
CACAGTG-------------------------ACAAAAACC
heptamer                         nonamer
```

Figure 18A

B.

```
                                  inversion breakpoint1
        11640       11650       11660    |   11670       11680
           *           *           *     |       *           *
dystrophin  T TTATAACAGT TATAAAGAAA GA^TTGTAAAC TAAAGTGTGC
            A AATATTGTCA ATATTTCTTT CT^AACATTTG ATTTCACACG
                                       a
apo-4 cDNA    840         850          |       870
[  138 ]    T TTATAACAGT TATAAAGAAA GA^TTaTAAAg gAAAaaGaaa>
            ^ ^^^^^^^^^^ ^^^^^^^^^^ ^^ ^^v^^^^v v^^^vv^vvv
dystrophin  T TTATAACAGT TATAAAGAAA GA^TTGTAAAC TAAAGTGTGC

-------------------------------------------------------------------------

                   11690       11700       11710       11720       11730
                     *           *           *           *           *
dystrophin  TTTATAAAAA AAAGTTGTTT ATAAAAACCC CTAAAAACAA AACAAACACA
            AAATATTTTT TTTCAACAAA TATTTTTGGG GATTTTTGTT TTGTTTGTGT
apo-4 cDNA   880    890         900       910         920         930
[  138 ]    aTaAaAtggA cAAGTgGTga ATgtgAACtC aggtgtgCAc AAttAtCAgg>
            v^v^v^vvv^ v^^^^v^^vv ^^vvv^^^v^ vvvvvvv^^v ^^vv^v^^vv
dystrophin  TTTATAAAAA AAAGTTGTTT ATAAAAACCC CTAAAAACAA AACAAACACA

-------------------------------------------------------------------------

                   11740     11750
                     *         *
dystrophin  CACACACACA CATACACACA
            GTGTGTGTGT GTATGTGTGT
apo-4 cDNA    940         950
[  138 ]    aACAC-CcCA -AaAC-CAaA>
            v^^^^ ^v^^  ^v^^ ^^v^
dystrophin  CACACACACA CATACACACA
```

Figure 18B

```
                                                             inv rsion breakpoint2
                   13130       13140       13150       13160   |   13170
                       *           *           *           *   |       *
dystrophin  AATTAGCTTT TGGAGAGTGG GTTTTGTCCA TTATTAATAA  TTAATTAATT
            TTAATCGAAA ACCTCTCACC CAAAACAGGT AATAATTATT  AATTAATTAA

                                                                     990
apo-4        ___________________________________________<AATTAATT
                                                          ^^^^^^^^
dystrophin                                                AATTAATT

------------------------------------------------------------------------
                    13180       13190       13200       13210       13220
                        *           *           *           *           *
dystrophin   AACATCAAAC ACGGCTTCTC ATGCTATTTC TACCTCACTT TGGTTTTGGG
             TTGTAGTTTG TGCCGAAGAG TACGATAAAG ATGGAGTGAA ACCAAAACCC

                   980         970         960         950         940
apo-4       <AACATCAAAC ACGGCTTCTC ATGCTATTTC TACCTCACTT TGGTTTTGGG
             ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^
dystrophin  AACATCAAAC ACGGCTTCTC ATGCTATTTC TACCTCACTT TGGTTTTGGG

------------------------------------------------------------------------
                   13230       13240       13250       13260       13270
                       *           *           *           *           *
dystrophin  GTGTTCCTGA TAATTGTGCA CACCTGAGTT CACAGCTTCA CCACTTGTCC
            CACAAGGACT ATTAACACGT GTGGACTCAA GTGTCGAAGT GGTGAACAGG

                   930         920         910         900         890
apo-4       <GTGTTCCTGA TAATTGTGCA CACCTGAGTT CACAGCTTCA CCACTTGTCC
             ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^
dystrophin  GTGTTCCTGA TAATTGTGCA CACCTGAGTT CACAGCTTCA CCACTTGTCC

------------------------------------------------------------------------

                   13280       13290       13300       13310       13320
                       *           *           *           *           *
dystrophin  ATTGCGTTAT TTTCTTTTTC CTTTATAATT CTTTCTTTTT CCTTCATAAT
            TAACGCAATA AAAGAAAAAG GAAATATTAA GAAAGAAAAA GGAAGTATTA
                                            |
                                  inversion breakpoint3
                   880         870         860         850         840
apo-4       <ATTGCGTTAT TTTCTTTTTC CTTTATAATT CTTTCTTTaT aacTgtTAta
         ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^v^ vvv^vv^^vv
dystrophin  ATTGCGTTAT TTTCTTTTTC CTTTATAATT CTTTCTTTTT CCTTCATAAT
```

Figure 18C

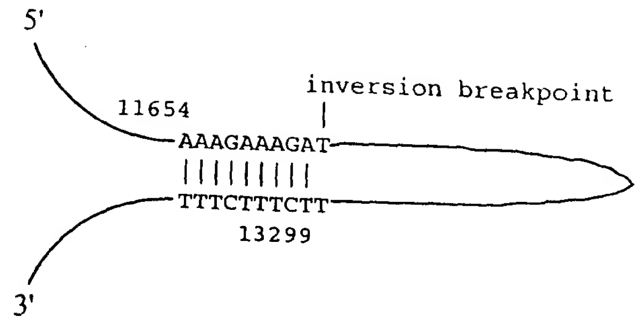
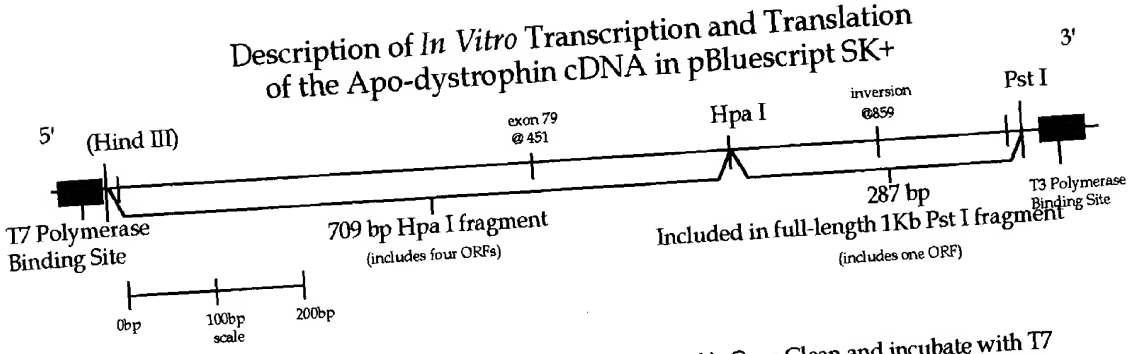


Figure 18D

```
                                inversion @ 860
    TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
841 ---------+---------+---------+---------+---------+---------+ 900
    ATTGTCAATATTTCTTTCTTAATATTTCCTTTTTCTTTTATTGCGTTACCTGTTCACCAC
```

Figure 19



Linearize plasmid with either Hpa I (truncated) or Pst I (full length). Gene Clean and incubate with T7 polymerase and dNTPs to produce RNA in vitro.

↓

Incubate RNA with Wheat Germ Extract or Rabbit Reticulocyte Lysates to produce *in vitro* translation

↓

Separate translation products by SDS-PAGE. Fix, Amplify and Dry Gel. Perform Autoradiography

Figure 20

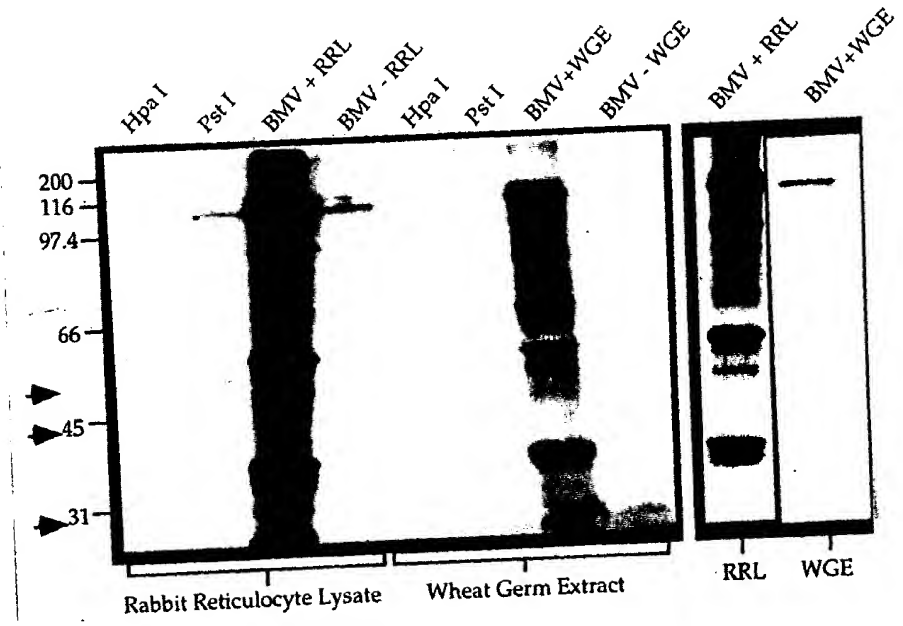


Figure 20A

Figure 20B

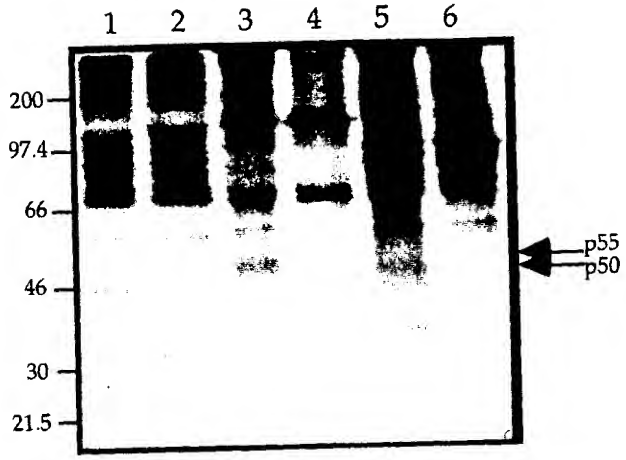


Figure 21

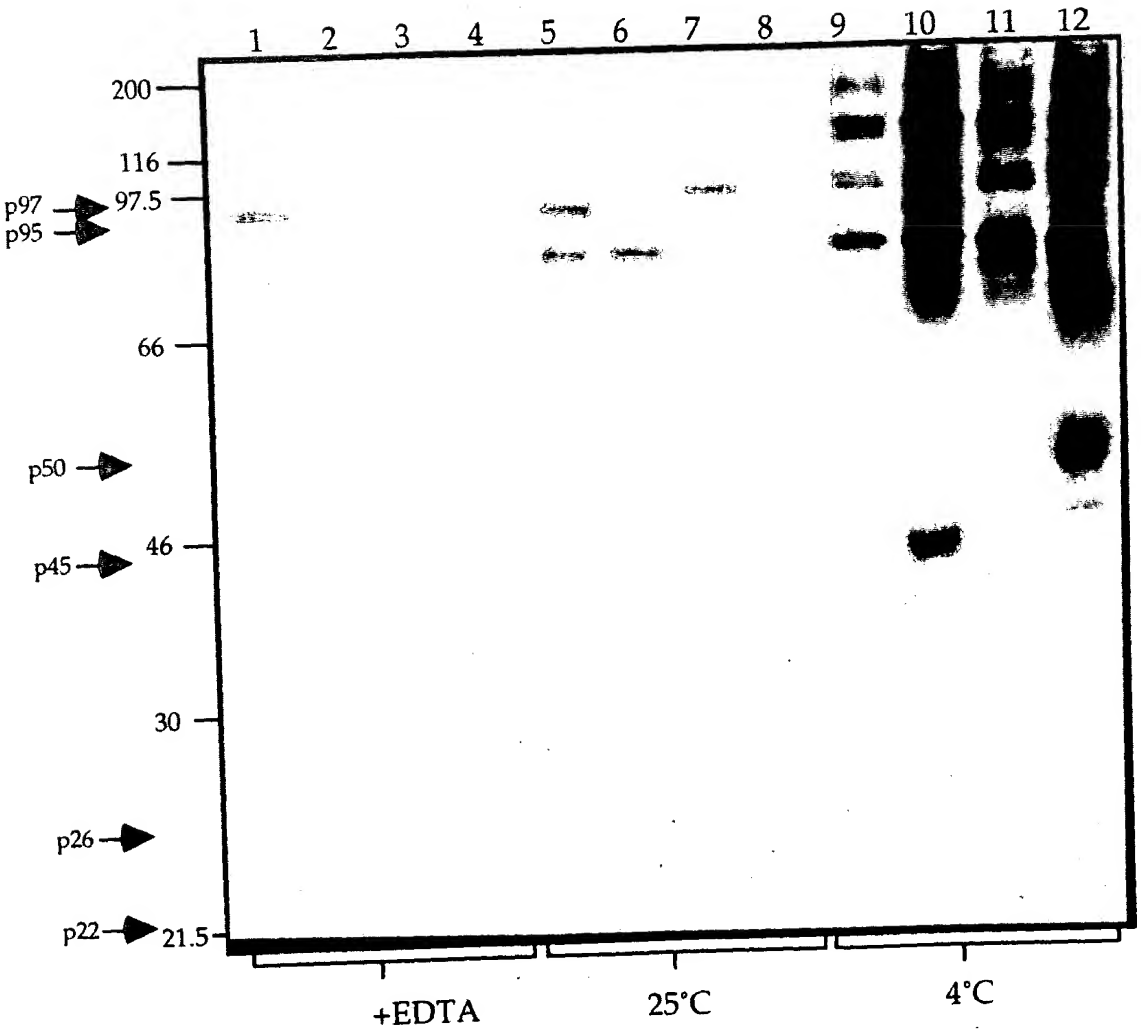


Figure 22

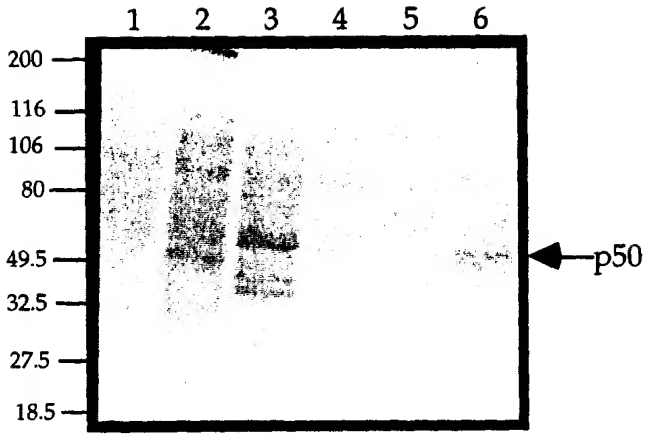


Figure 23

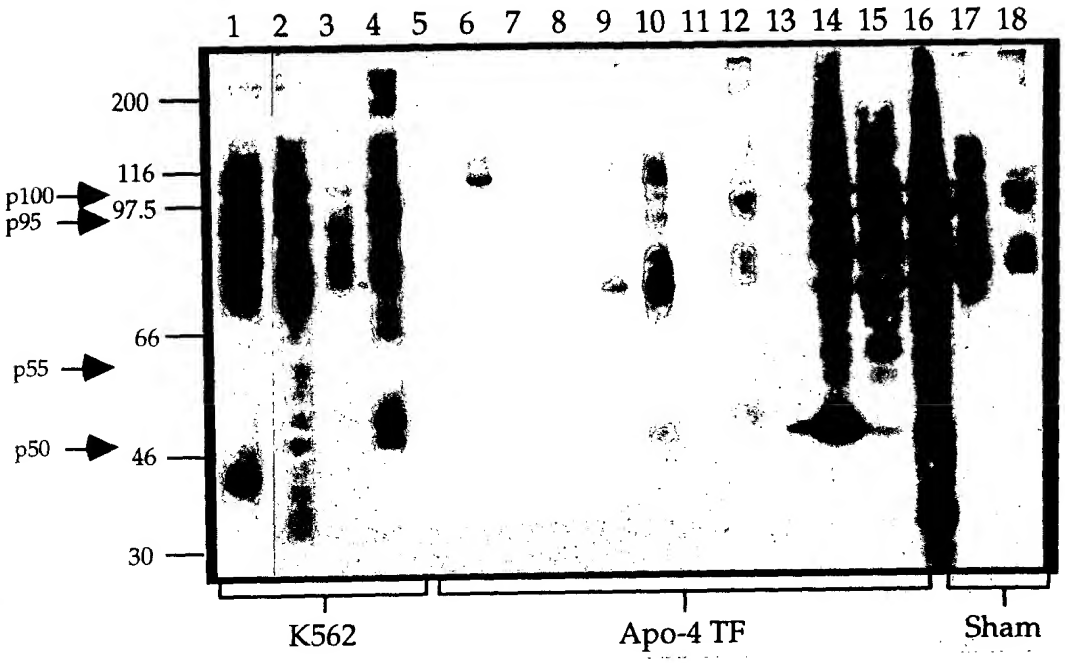


Figure 24

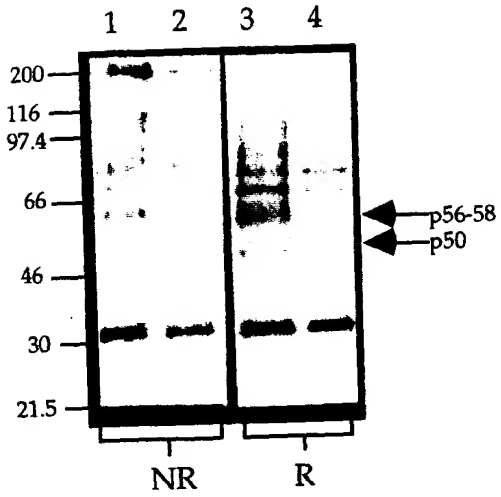


Figure 25A

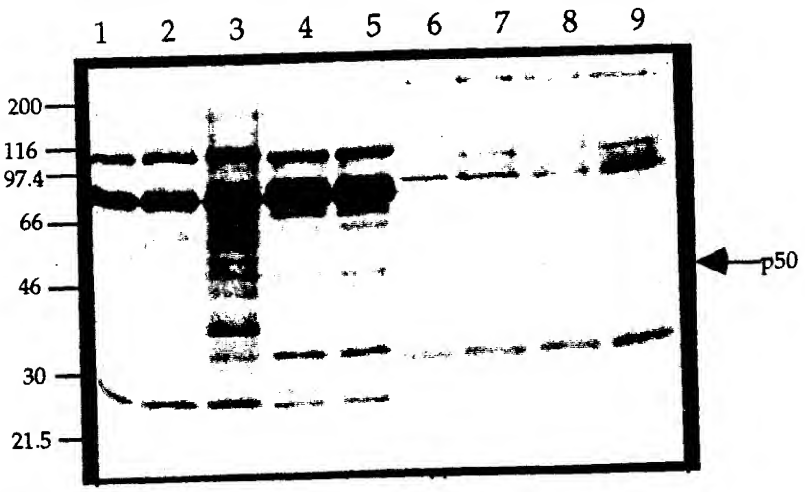


Figure 25B

H2 starting at second methionine - 321 bp, predicted weight = 17.4Kd + 1 N-glycosylation site + 20.4 Kd.

Figure 26A

Splice sites for peptide
MYPIMEYSCSDRNLVLIYGILLIYIYIGNLNMKKEQNKCFTTPDSRMVFIIFIQQRGLDSKSLQEINL
YFCEGFYTSMQLYKKVIRKLHKITQWTRTPQNQSEVEIA 107

Figure 26B

Start	Exon No.	Exon Position	Exon Length	Intron No.	Intron Position	Intron Length
@88 bp	78.3	@74-180	106 bp	79.1	@181-529	349 bp
	79.1	@530-654	125 bp	79.4	@655-720	66 bp
	79.4	@721-769	49 bp	79.55	@770-875	105 bp
	79.55	@876-893	18 bp	79.75	@894-932	39 bp
	79.85	@933-966	33 bp			

Hydrophobicity Scale KD; Candidate membrane-spanning segments:

Certain 1 12- 32 1.8833

Figure 26C

Predicted TM structure

- > : Too long to be significative
- < : Too short to be significative
- Ll : Loop length
- KR : Number of Lys and Arg
- KR Diff : Positive charge difference
- CE : Net charge energy
- CE Diff : Net charge difference
- CH Diff : Charge difference over N-term segments

KR Diff = 1 CYTOPLASM

CH Diff = -3 OUTSIDE

CE Diff = 0.54 OUTSIDE

Structure no. 1

Figure 26D

A readthrough apo-4S product using the second available methionine

The Apo-4S peptide sequence

```
P1              Begin TM1(R)
+30             |             P2
MYPIMEYSCSD RNLVLIYGIL LIYIYIGNLN VARHFSMKTP VARSNIKLIL 80

TNNVKWLHKK GFASSWKLVK NQTLLCTPSM QLLCCLHPEM GNDFPNGKET 130
                                          P3
ERCYLSAPFV KSVFLSLCFP GHNVGSLFHM ADDLGRAMES LVSVMTDEEG 180

AEKMFYNSRF PHGFYNIHTT KRIRQKEFTR NKSIFLRRVV VLYCRFQKFL 230

SLLLFCKQWQ VLHVYAIVQK SYKKTTCKIL IAKKLAISLY GTHFGLFKNL 280

KQLKRKNYKG KRKKRNGQVV KLRTQVCTII RNTPKPKRGR NSMRSRVRCK 330

LI 332 (302aa in predicted polypeptide)
```

Figure 27A

Candidate membrane-spanning segments:

Certain	1	41-61	1.9073
Putative	2	101-121	0.8052
Certain	3	132-152	1.2552
Putative	4	217-237	1.1833
Putative	5	254-274	0.9240

Transmembrane segments included in structure No. 8: 1 2 3 4 5

Loop lengths: 11 39 10 64 16 58; K+R profile: 1 2 5 (9 >22)

K+R difference: -23: -> Orientation: **N-out**

Charge-difference over N-terminal Membr. segs. (±15 residues): -4
-> Orientation: **N-out**

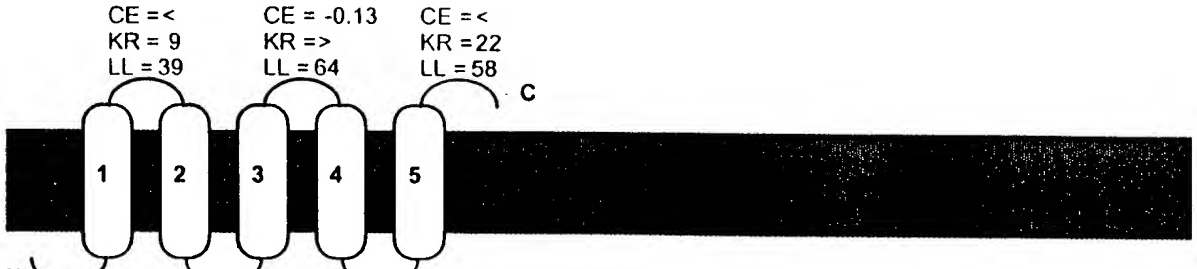
CYT-EXT profile (neg. values indicate cytoplasmic preference): < -0.13 <

CYT-EXT difference: 0.13: -> Orientation: **N-out**

Figure 27B

> : Too long to be significative
< : Too short to be significative
LI : Loop length
KR : Number of Lys and Arg

KR Diff : Positive charge difference
CE : Net charge energy
CE Diff : Net charge difference
CH Diff : Charge difference over N-term segments



KR Diff = -23 OUTSIDE
CH Diff = -4 OUTSIDE
CE Diff = 0.13 OUTSIDE

Structure no. 8

Figure 27C

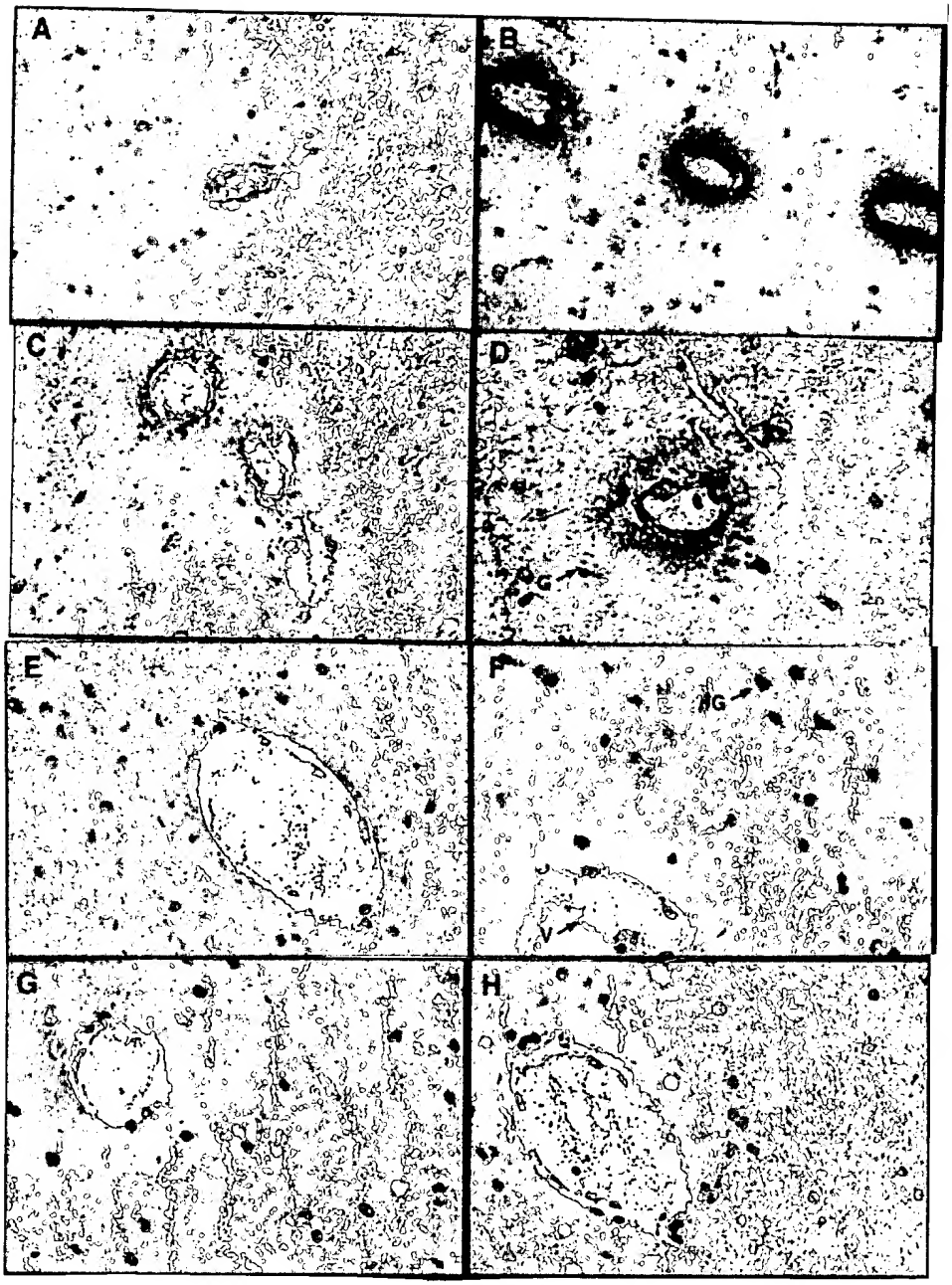


Figure 28

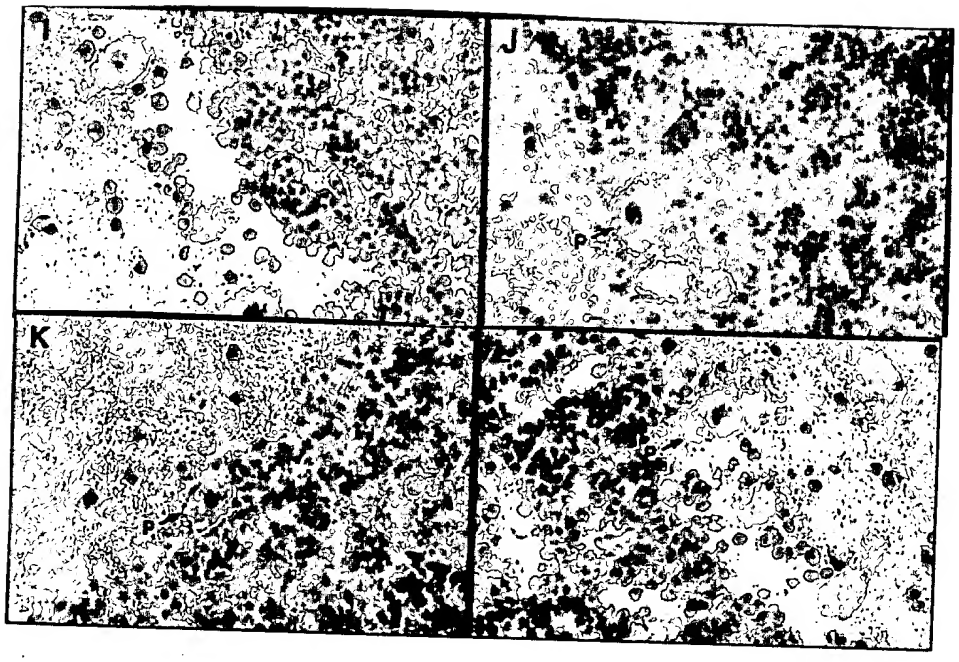


Figure 28 (cont'd)

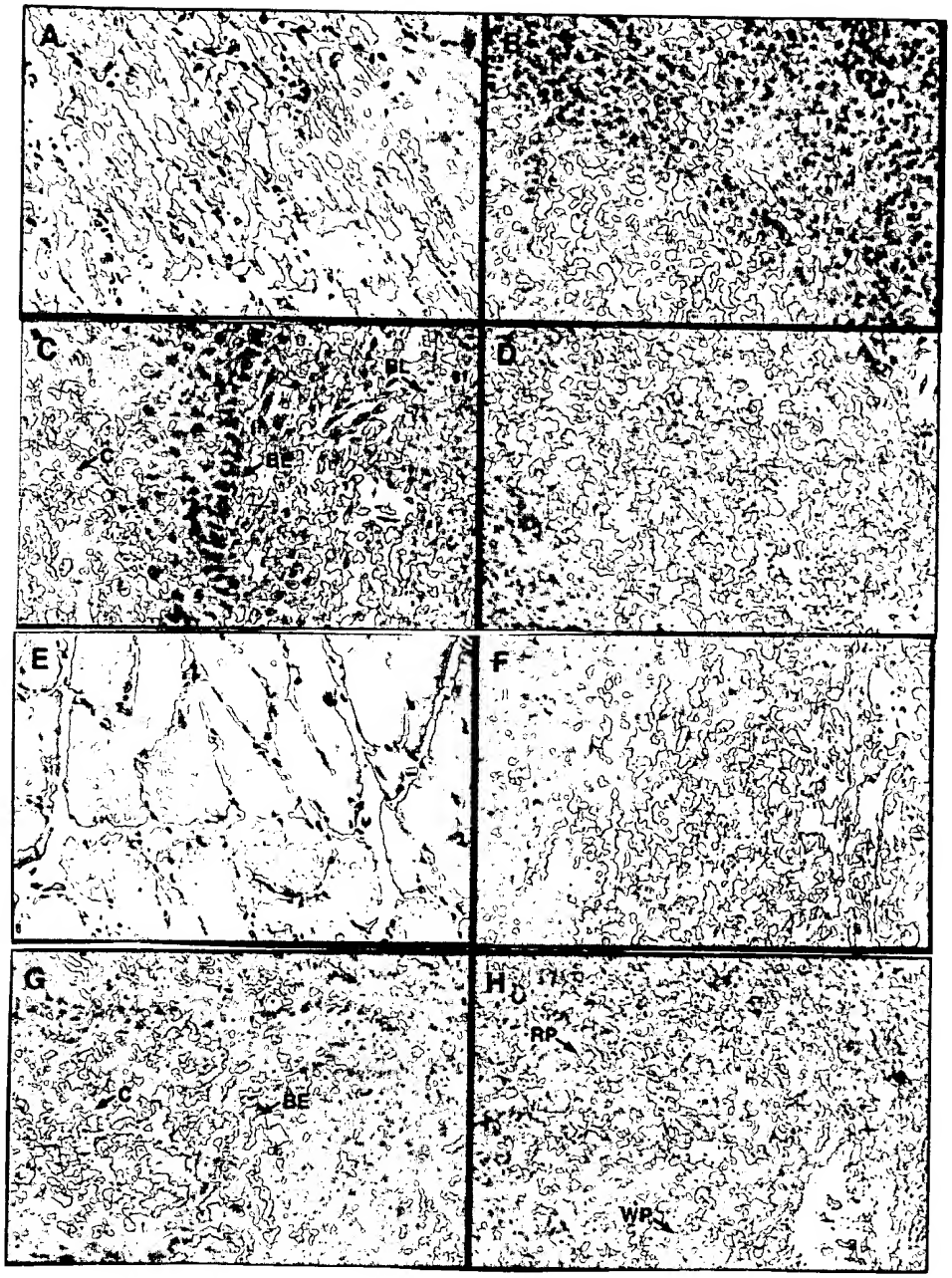


Figure 29

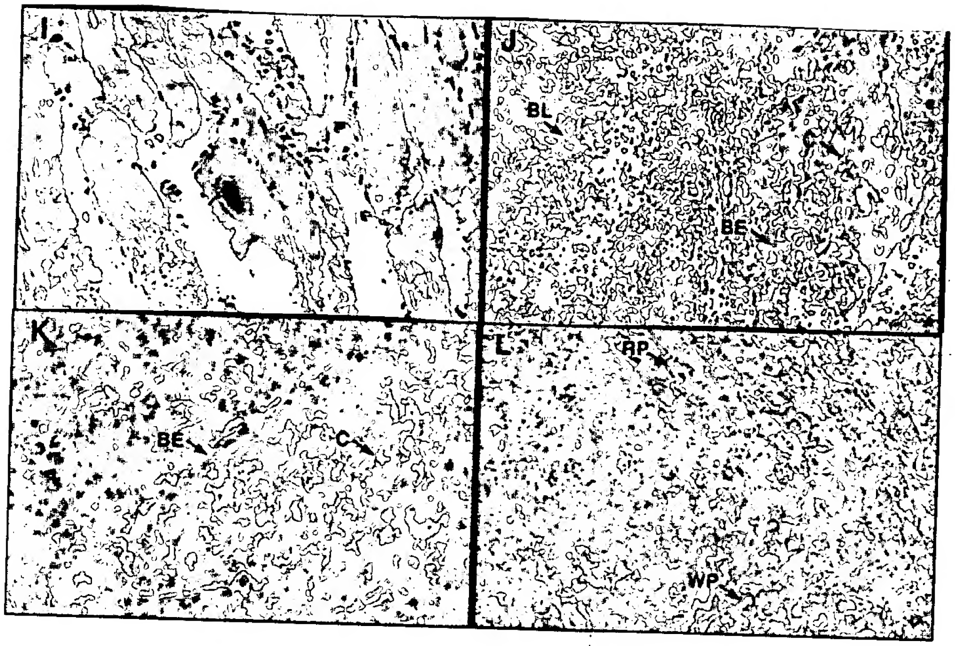


Figure 29 (cont'd)

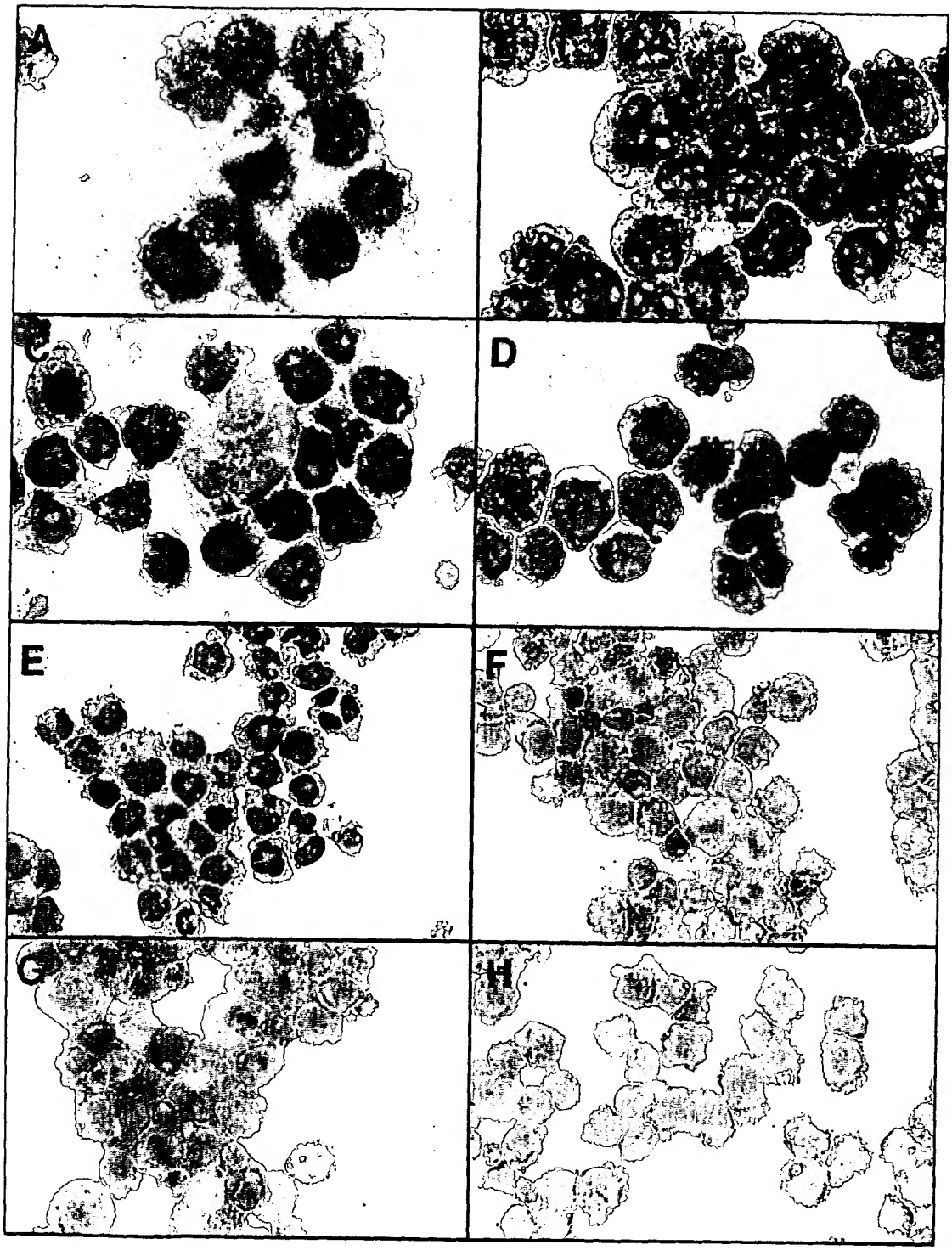


Figure 30

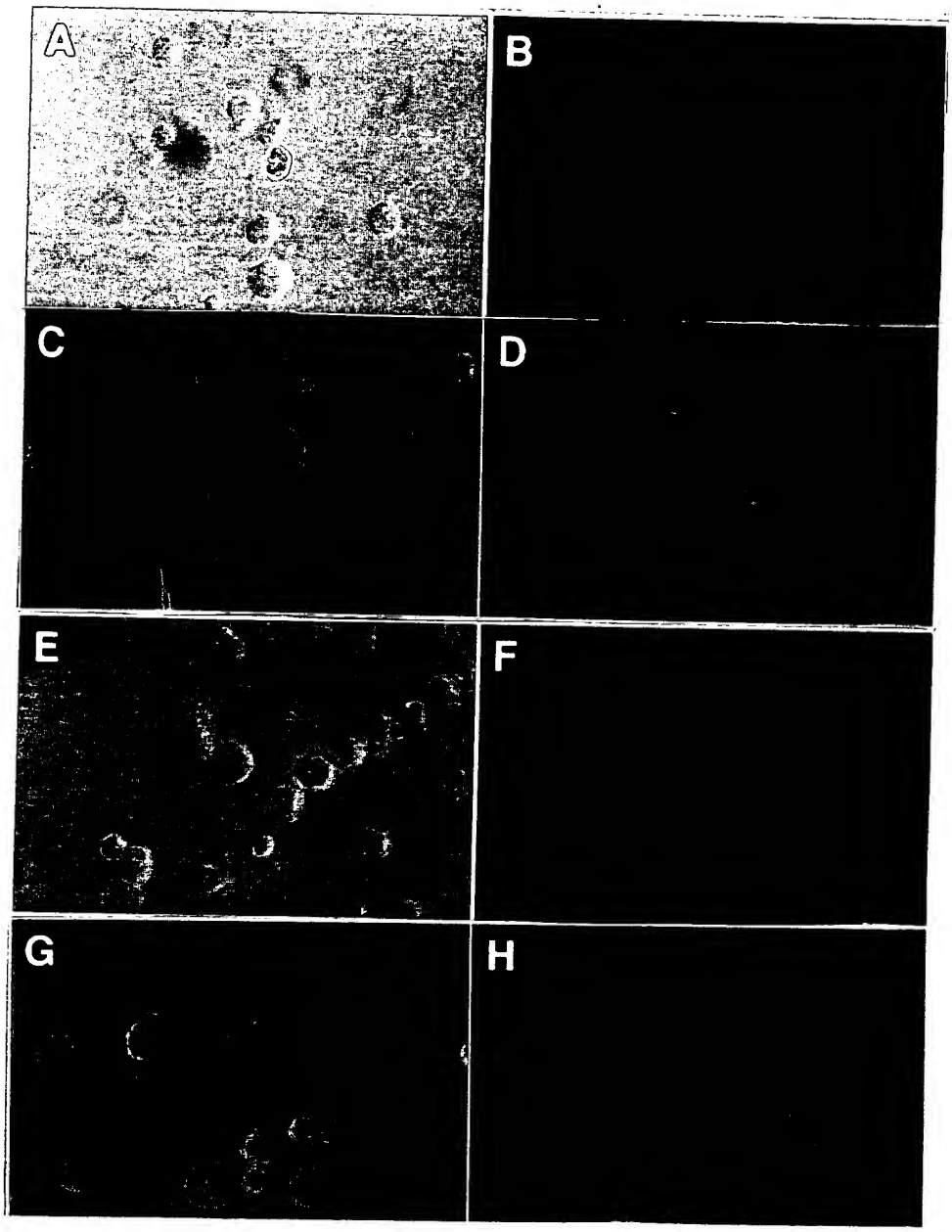


Figure 31

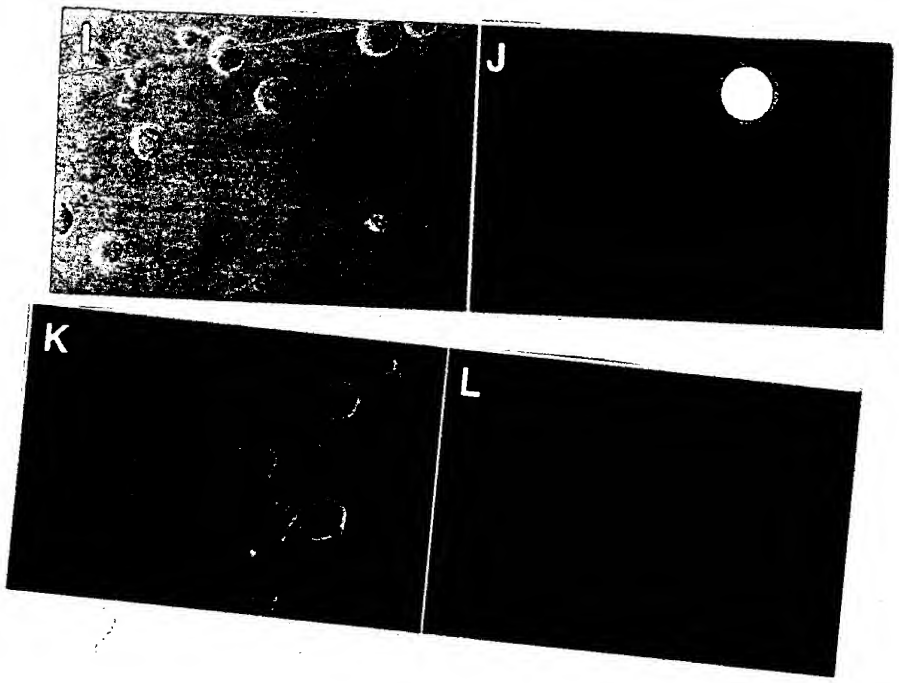


Figure 31 (cont'd)

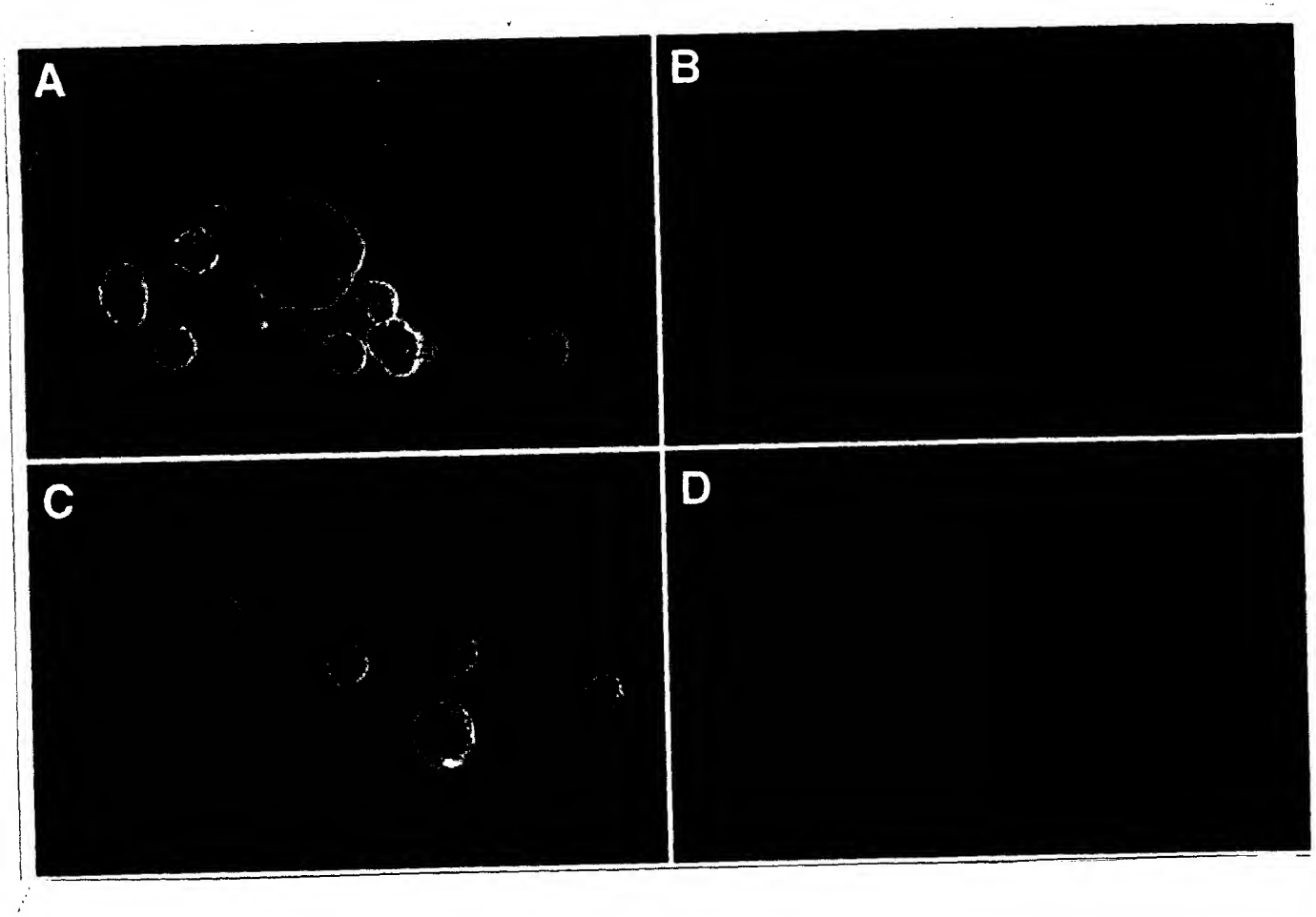


Figure 32

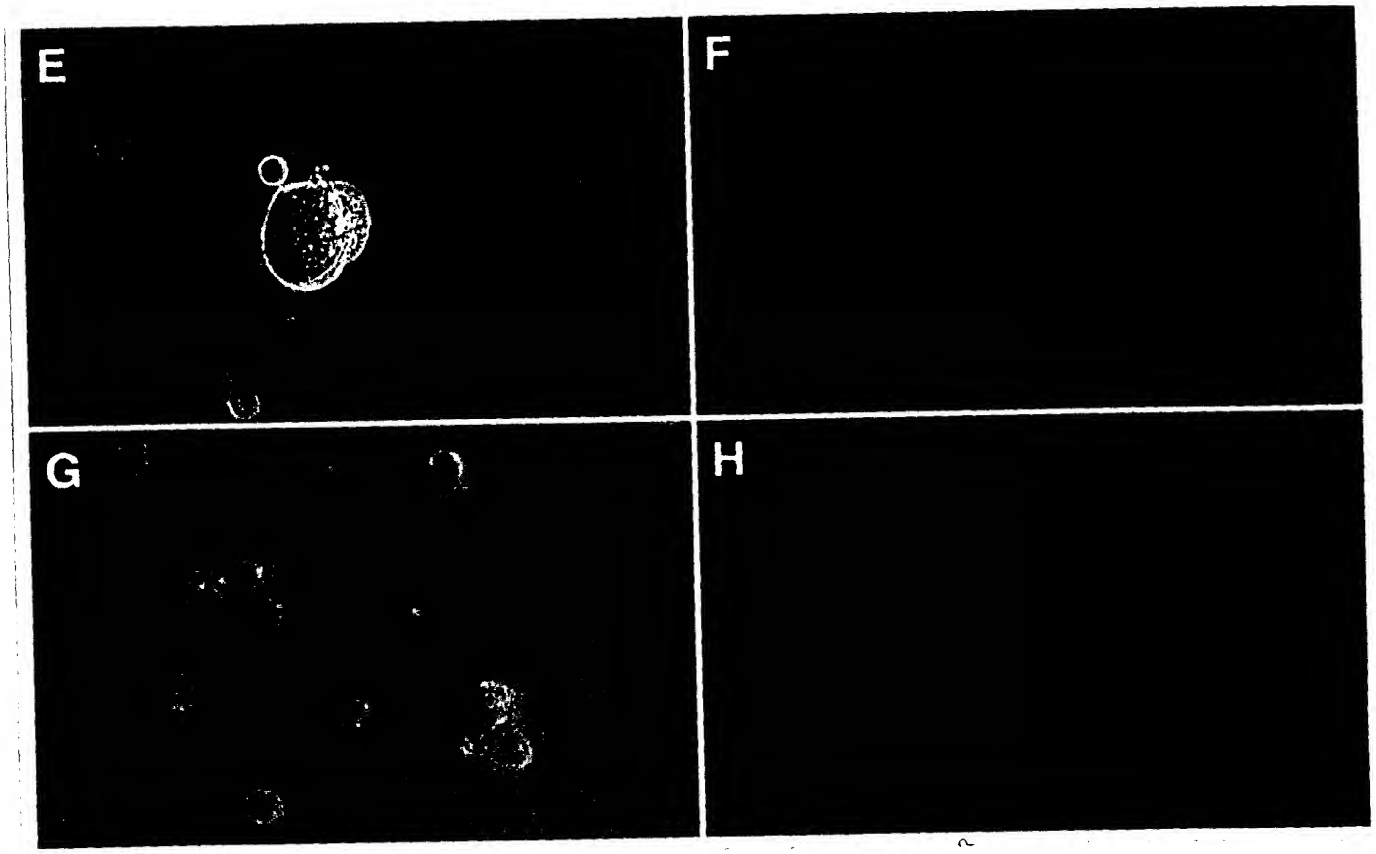


Figure 32 (cont'd)

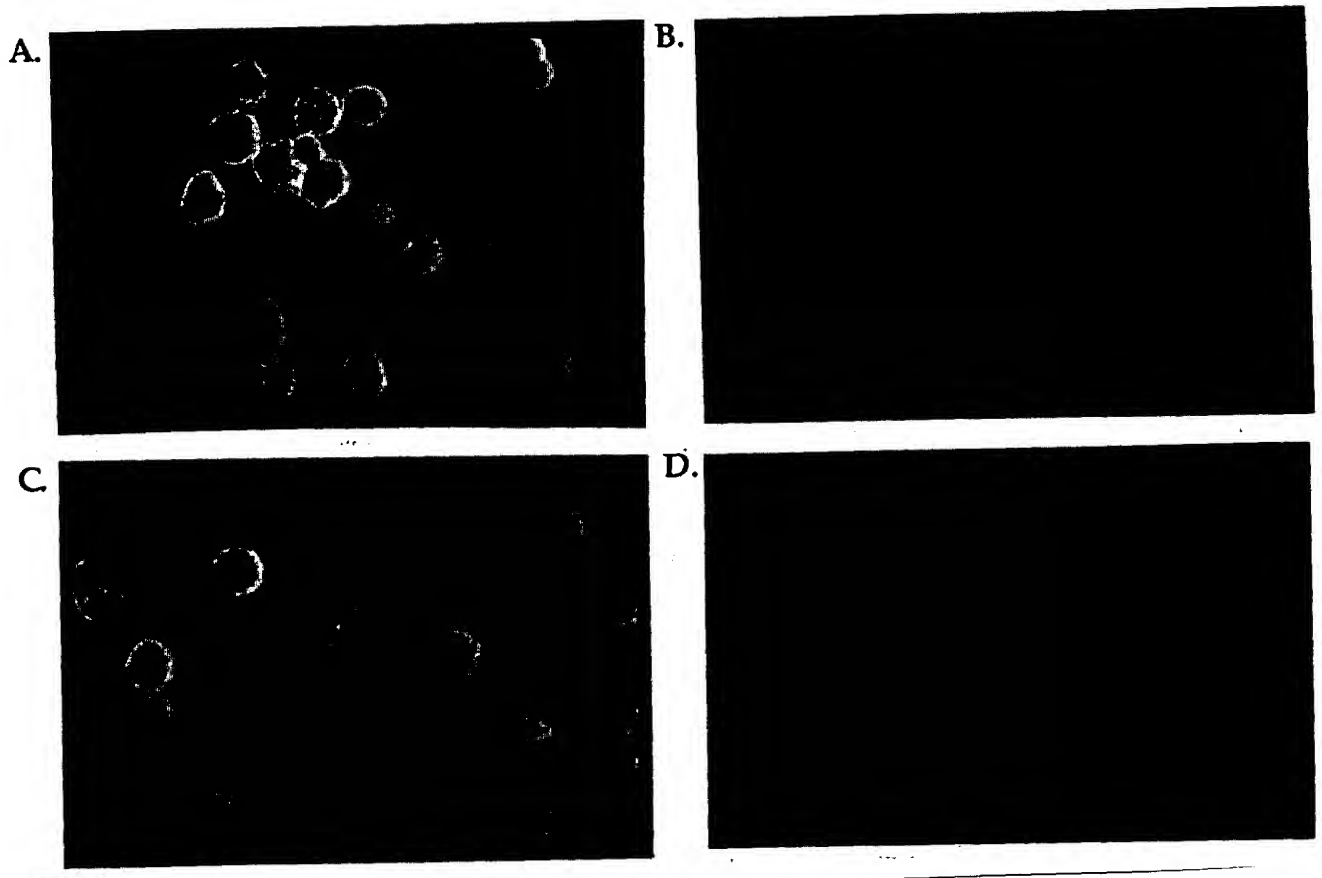


Figure 33

Additional Oligonucleotide primers used for apo-dystrophin-4 southern blotting and sequencing

FORWARD

GTT CGT TAA TAC AAG TAG	F2.3(@28)	(SEQ ID NO 15)
GCC AAG GTG GAA AAG ATG	F2.2(@73)	(SEQ ID NO 16)
CCA GTA GCC TGA TCC AAC	F3.2(@208)	(SEQ ID NO 17)
GGC TTC ATT AAT AAG	F3.1(@257)	(SEQ ID NO 18)
GGC AAA GAA ACA GAG TG	F4.2(@379)	(SEQ ID NO 19)
CAG GAC ACA ATG TAG GA	F4.1(@449)	(SEQ ID NO 20)
GTT ATA AAG AAA GAA TTA TAA AG	FJn(@846)	(SEQ ID NO 21)
GAA AAT AAC GCA ATG GAC	F5.1(@875)	(SEQ ID NO 22)

REVERSE

GAT GGG ATA CAT CTT TTC C	R6.1(@99)	(SEQ ID NO 23)
CAA GCT ACA TTC AGG TTC CC	F2.2R(@188)	(SEQ ID NO 24)
GGA CTC CAT CGC TCT GCC	R4.1(@510)	(SEQ ID NO 25)
GAC TTA GAA ACT ACT G	R3.4(@694)	(SEQ ID NO 26)
ATA GAC GTG TAA AAC CTG C	R2.1(@735)	(SEQ ID NO 27)
AAC TGT TAT AAA TTT TTA	RSP2(@848)	(SEQ ID NO 28)
CTT TTT CCT TTA TAA TTC TTT C	R2.3o(@875)	(SEQ ID NO 29)

Figure 34

An Additional Splice Product Predicted From The Apo-4 Gene

A second potential theoretical splice product which retains exon 78.3 is shown below.

H2 p1-124 spliced product =351 bp, 117 amino acids + 10 from vector + 1 N-glycosylation site; predicted weight = 21.9 Kd

Figure 35A

Peptide Generated

MFVNTTKVEKMYPIMEYSCSDRNLVLIYGILLIYIYIGNLNMKKEQNKCFTTPDSRMVFII FIQQRGLDSKSLQEINLYFCEGFYTSMQLYKKVIRKLHKITQWTRTPQNQSEVEIA (117 amino acids) (SEQ ID NO 30)

Figure 35B

Start	Exon No.	Exon Position	Exon Length	Intron No.	Intron Position	Intron Length
@26 bp	78.1	@16-41	26 bp	78.3	@42-74	33 bp
	78.3	@75-181	106 bp	79.1	@182-530	349 bp
	79.1	@531-655	125 bp	79.4	@656-721	66 bp
	79.4	@722-770	49 bp	79.55	@771-876	105 bp
	79.55	@877-894	18 bp	79.75	@895-933	39 bp
	79.85	@934-967	33 bp			

Hydrophobicity Scale KD; Candidate membrane-spanning segments:

Certain 1 22- 42 1.8833

Figure 35C

Predicted TM structure

\>: Too long to be significative
<: Too short to be significative
U: Loop length
KR: Number of Lys and Arg

KR Diff: Positive charge difference
CE: Net charge energy
CE Diff: Net charge difference
CH Diff: Charge difference over N-term segments

KR Diff = 3 CYTOPLASM
CH Diff = -2 OUTSIDE
CE Diff = 0.54 OUTSIDE

Structure no. 1

Figure 35D

Nucleic Acid Subsequence Sites Identified In Apo-4

Motif	Position	Significance
CpG	-7, (+28, +106)	DNA methylation site
CAAT	-132, (+127, +131)	Binding of CAAT factors
TATAAT (5/6)	-120, -114, (+10)	TFIID Binding site
TATA	-154	Binds RNA polymerase II and TFIID
CCATTCA	-162, -131	Cap Site I
TATCAGT	+12, (+25)	Cap Site II
TGGCTGCAAGCCCAA (10/14)	-57, (+41)	Binds CTF/NF-I protein
GTGATGG	-140, -4, +11, +32	Eucaryotic Transcription Initiation Site

Figure 36

Top Pred predicts 4-5 transmembrane domains for a full-length apo-4F product in which all the stop codons are suppressed.

Protein sequence and position of predicted TM domains

```
                                   Begin TM1(R)
                          P1              |            P2
MFVNTSREKV INQSLIAKVE KMYPIMEYSCSD RNLVLIYGIL LIYIYIGNLN VARHFSMK60

TPVARSNIKL ILTNNVKWLH KKGFASSWKL VKNQTLLCTP SMQLLCCLHP EMGNDFPNGK 120
                                         P3
ETERCYLSAP FVKSVFLSLC FPGHNVGSLF HMADDLGRAM ESLVSVMTDE EGAEKMFYNS180

RFPHGFYNIH TTKRIRQKEF TRNKSIFLRR VVVLYCRFQK FLSLLLFCKQ WQVLHVYAIV 240

QKSYKKTTCK ILIAKKLAIS LYGTHFGLFK NLKQLKRKNY KGKRKKRNGQ VVKLRTQVCT 300

IIRNTPKPKR GRNSMRSRVR CKLI (324 amino acids) (SEQ ID NO 31)
```

Hydrophobicity Scale KD

Figure 37A

Apo-4F : Candidate membrane-spanning segments:

Certain	1	33- 53	1.9073
Putative	2	93- 113	0.8052
Certain	3	124- 144	1.2552
Putative	4	209- 229	1.1833
Putative	5	246- 266	0.9240

I. Transmembrane segments included in structure 8: **1 2 3 4 5;** Loop lengths: 32 39 10 64 16 58

Figure 37B

K+R difference: -19; -> Orientation: **N-out**; Charge-difference over N-terminal Membr. segs.

(±15 residues): -3; -> Orientation: **N-out**

CYT-EXT profile (neg. values indicate cytoplasmic preference): < < < < -0.13 <

CYT-EXT difference: 0.13

-> Orientation: **N-out**

II. Transmembrane segments included in structure 7: **1 3 4 5;** Loop lengths: 32 70 64 16 58

K+R profile: 5 > 22 > 5; K+R difference: 22 -> Orientation: **N-in**

Charge-difference over N-terminal Membr. segs. (±15 residues): -3; -> Orientation: **N-out**

CYT-EXT profile (neg. values indicate cytoplasmic preference): < -0.13 < -0.26 <

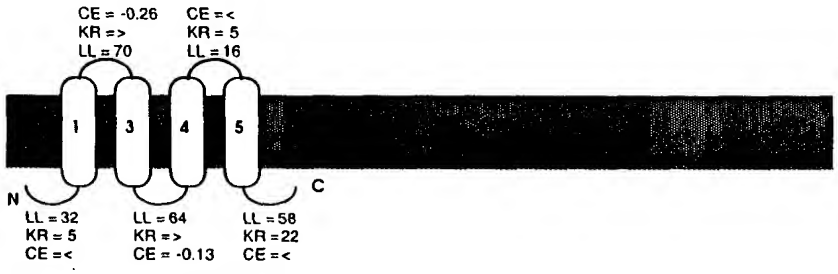
CYT-EXT difference: 0.13; -> Orientation: **N-out**

Figure 37B (cont'd)

TopPred predicts a cytoplasmic N-terminus for four TM domains

> : Too long to be significative
< : Too short to be significative
LI : Loop length
KR : Number of Lys and Arg

KR Diff : Positive charge difference
CE : Net charge energy
CE Diff : Net charge difference
CH Diff : Charge difference over N-term segments



Structure no. 7

Figure 37C

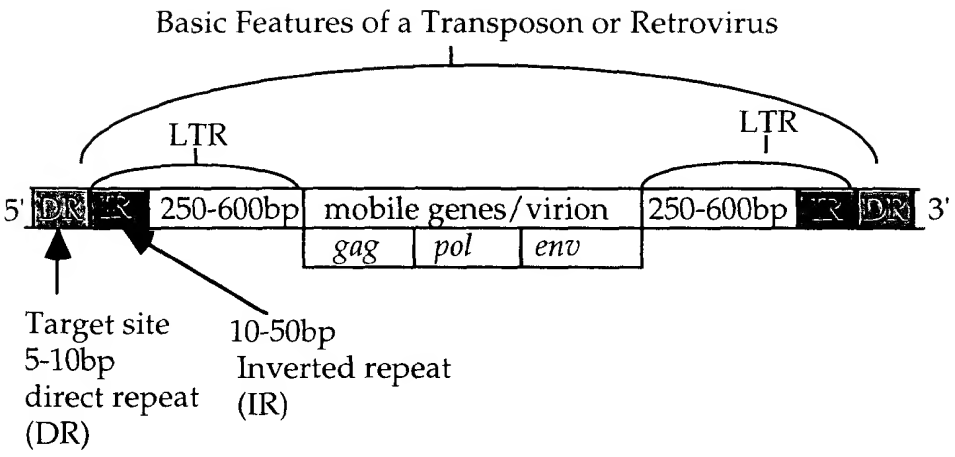


Figure 38A

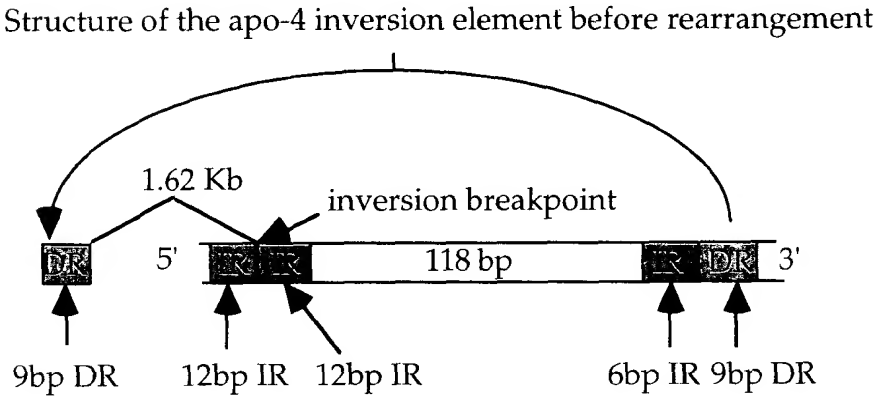
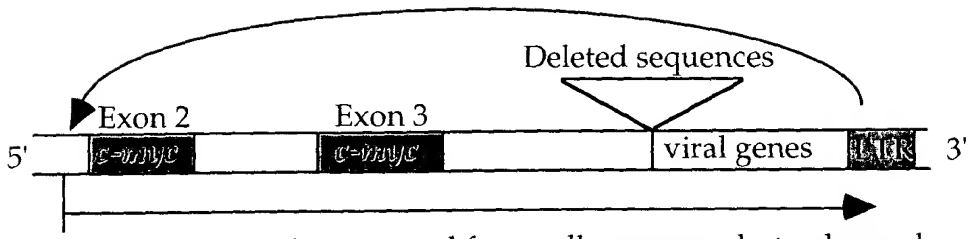
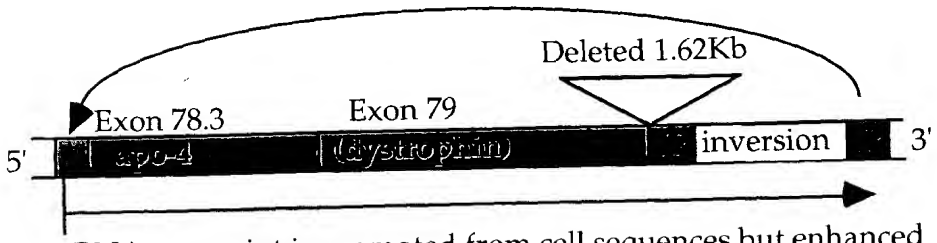


Figure 38B



RNA transcript is promoted from cell sequences but enhanced and terminated by viral sequences.

Figure 39A



RNA transcript is promoted from cell sequences but enhanced and terminated by inversion sequences which may also activate suppressor tRNAs or reverse transcriptase activity to prevent the recognition of stop codons. Inverted repeats (IR) are present at both ends of the inversion, as they are in retroviruses and transposable elements.

Figure 39B